Econometric Theory, **35**, 2019, 842–899.
doi:10.1017/S0266466618000294

ASYMPTOTICALLY EFFICIENT MODEL SELECTION FOR PANEL DATA FORECASTING

RYAN GREENAWAY-MCGREVY
The University of Auckland

This article develops new model selection methods for forecasting panel data using a set of least squares (LS) vector autoregressions. Model selection is based on minimizing the estimated quadratic forecast risk among candidate models. We provide conditions under which the selection criterion is asymptotically efficient in the sense of Shibata (1980) as n (cross sections) and T (time series) approach infinity. Relative to extant selection criteria, this criterion places a heavier penalty on model dimensionality in order to account for the effects of parameterized forms of cross sectional heterogeneity (such as fixed effects) on forecast loss. We also extend the analysis to bias-corrected least squares, showing that significant reductions in forecast risk can be achieved.

1. INTRODUCTION

There are significant advantages to using panel data models to forecast economic variables. In empirical forecasting applications, panel models that permit limited cross sectional heterogeneity typically produce more accurate forecasts than the corresponding time series specification. This finding holds across a wide range of macroeconomic and microeconomic data, including panel datasets with a large time series dimension.[1] For a detailed survey, see the "Forecasting Applications" section of Baltagi (2008).

Despite the empirical advantages of forecasting with panel data, there is little theory to guide the practitioner when choosing the specification of a panel forecasting model. Much of the extant theoretical research on panel forecasting

The author thanks Yong Bao, Peter C.B. Phillips, Donggyu Sul, and seminar participants at the University of Auckland, the 23rd NZESG meeting, and the 22nd Midwest Econometrics Group meeting for their comments. This work was supported in part by the Marsden Fund Council from Government funding, administered by the Royal Society of New Zealand under grant No. 16-UOA-239. Address correspondence to Ryan Greenaway-McGrevy, Department of Economics, The University of Auckland, Auckland, New Zealand; e-mail: r.mcgrevy@auckland.ac.nz.

[1] For forecasting exchange rates, see Mark and Sul (2001) and Rapach and Wohar (2004); for migration, see Brucker and Siliverstovs (2006); for gasoline consumption, see Baltagi and Griffin (1997); for electricity consumption, see Baltagi, Bresson, and Pirotte (2002); for cigarette consumption, see Baltagi, Griffin, and Xiong (2000); for firm-level investment, see Driver, Imai, Temple, and Urga (2004); for output, see Hoogstrate, Palm, and Pfann (2000); for a VAR forecast of key macroeconomic variables, see Gavin and Theodorou (2005).

focuses on efficient prediction for a given model specification (see, among others, Taub, 1979; Baltagi and Li, 1992; Baillie and Baltagi, 1999; Baltagi, Bresson, and Pirotte, 2010). This literature also assumes strict exogeneity of the regressor, which precludes models with lagged dependent variables. By contrast, research focusing on the relationship between model specification and out-of-sample loss has been limited to a few articles (e.g., Greenaway-McGrevy, 2013; 2015).

In this article we provide new model selection criteria that are explicitly designed to choose a panel data model for the purpose of out-of-sample forecasting. We consider a set of panel vector autoregressions (VARs) that permit limited forms of parameterized heterogeneity, including cross section fixed effects or heterogenous linear trends. Dependence within each time series is modelled through a homogenous vector autoregressive structure that can be estimated by pooling across the time series in the sample. See (4) below for the model specifications considered.

The proposed model selection methods are based on minimizing expected quadratic forecast loss among candidate models. The estimated models are used to make a one-period-ahead out-of-sample forecast of each cross sectional unit in the panel. Forecast loss is then evaluated using an average quadratic forecast error loss function, with the average taken over the cross sectional units of the panel. We construct an estimator of the expected quadratic forecast loss so that model selection can proceed by choosing the model with the lowest estimated forecast risk. We then provide conditions under which this selection criterion is *asymptotically efficient* in the sense defined by Shibata (1980). Asymptotically efficient selection criteria minimize the expected quadratic forecast loss within the candidate set of models as the sample size grows large.

In studying the relationship between model specification and the out-of-sample loss, we will adhere to a similar framework to that set out in the theoretical time series forecasting literature. Akaike (1970), Shibata (1980), Fuller and Hasza (1981), Findley (1984), Kunitomo and Yamamoto (1985), Bhansali (1996, 1997), Ing (2003), Ing and Wei (2003, 2005) and Schorfheide (2005), among others, derive expressions for the expected quadratic forecast loss of a least squares (LS) fitting to a set of covariance stationary time series processes. These expressions form the basis for deriving and evaluating well-known forecasting model selection methods, such as minimization of Akaike's (1970) final prediction error (FPE), Shibata's (1980) FPE, Mallows' (1973) C_p, and the Akaike information criterion (Akaike, 1970; Shibata, 1980; Bhansali, 1996; Ing and Wei, 2005). These selection criteria are typically evaluated within an asymptotic framework that preserves the trade-off between specification error and over-parameterization that we face in practice when selecting the size of a model (e.g., Shibata, 1980; Speed and Yu, 1993; Bhansali, 1996; Ing and Wei, 2005; Schorfheide, 2005). Most commonly the data is generated by an infinite-dimensional specification of the class of models under consideration, and the set of candidate models is permitted to increase with the sample size at a restricted rate. For example, Shibata (1980), Bhansali (1996) and Ing and Wei (2003, 2005) consider fitting an AR(k) model to an AR(∞) process, permitting the set of fitted lag orders to grow with the sample size T at an $o(T^{1/2})$ rate. In this framework, all candidate models can exhibit some degree of misspecification, and the goal of model selection is to choose the model that can best balance misspecification against over-parameterization for the purpose at hand.

Following this precedent, we permit the vector autoregressive component of the panel to be of infinite order, and we allow the maximum number of lags defining the set of candidate models to grow at a restricted rate as both n (number of cross sections) and T (number of time series) approach infinity (see Assumption 3 below). Under this framework we derive an analytic expression that uniformly (across all models) approximates the expected quadratic forecast error loss function of the LS fitting (henceforth referred to as *quadratic forecast risk*, or QFR).

A key result is that the asymptotic QFR is increasing in both the variance and a quadratic transformation of the finite sample bias of the LS estimator. It is well-established that LS estimators of dynamic panel data models such as VARs exhibit $O\left(T^{-1}\right)$ bias (Nickell, 1981; Hahn and Kuersteiner, 2002; Alvarez and Arellano, 2003; Phillips and Sul, 2007; Lee, 2012), and a quadratic transformation of this bias is manifest in the QFR (also see Greenaway-McGrevy, 2013; 2015).[2] This result contrasts with the conventional result for covariance stationary time series, where only the variance of the LS estimator is manifest in the relevant approximation of the QFR.

The manifestation of the $O\left(T^{-1}\right)$ bias (or "Nickell bias") in the QFR implies that extant forecast selection criteria that penalize only the variance of the LS fitting, such as minimization of FPE (Akaike, 1970; Shibata, 1980) or Mallows' C_p (Mallows, 1973), are inappropriate in the panel data context. In addition, any cross sectional correlation in the panels must be accounted for when penalizing the variance of the LS fitting. We therefore use our findings to design new model selection methods that are tailored to panel data forecasting. Specifically, we augment a version of the conventional Akaike FPE with a quadratic bias term that is estimated using the fitted model parameters. To account for potential cross sectional correlation, we cluster by time period when estimating the variance of the LS fitting. Lag order selection then proceeds by choosing the model with the minimum estimated QFR, in the vein of the model selection methods suggested by Akaike (1969, 1970), Shibata (1980) and Bhansali (1996). We then show that the selection method is asymptotically efficient provided that one dimension of the panel (either n or T) grows at a slightly faster rate than the other dimension of the panel (see Theorem 3.2 for additional details).

Due to the effects of LS estimator bias, this new selection criterion places a heavier penalty on model complexity compared to conventional forecast selection criteria, such as the Akaike and Shibata versions of FPE. In this sense it is similar

[2] Our use of the term *bias* differs from how the term is typically used elsewhere in the forecasting literature, where it is commonly used to refer to the effect of misspecification on the forecast.

to the panel data selection criteria recently proposed by Lee and Phillips (2015), which also impose an additional penalty on model complexity in order to account for parameterized cross sectional heterogeneity. However, our panel FPE differs from the Lee and Phillips criteria in some other key respects. The latter criteria are designed to select a model that best approximates the data *after integrating out* nuisance parameters such as fixed effects from the likelihood function. The panel FPE selection criterion, on the other hand, imposes an additional penalty on model dimension in order to account for the impact of variability in the estimated heterogenous parameters on forecast loss. The panel FPE criterion therefore takes into account the impact of nuisance parameters such as fixed effects when selecting a model.

We also extend our analysis to bias-corrected (BC) least squares forecasting. Under this approach we use analytic expressions for the bias in conjunction with the fitted LS parameters to make a first order correction for the bias *before* making the forecast. We show that the asymptotic QFR of the BCLS fitting is strictly less than that of the LS fitting, revealing an asymptotic justification for the use of the method in forecasting applications. We provide an analogous model selection criterion that is based on minimization of the estimated QFR of the BCLS forecast, and we outline conditions for the asymptotic efficiency of the selection criterion.

The remainder of the article is organized as follows. In the following section we outline our assumptions and forecasting methods. In Section 3 we derive an asymptotic expression for the QFR of the LS forecast, illustrating how this expression informs the structure of the panel FPE. We then establish conditions under which model selection via minimization of panel FPE is asymptotically efficient, and provide analogous results for BCLS forecasts. In Section 4 we conduct Monte Carlo studies in order to validate the asymptotic theory and explore the small sample performance of the proposed model section methods. Section 5 concludes. Throughout, ":=" is used as the definitional equality; C denotes an arbitrary finite constant; $\mathrm{tr}(\cdot)$, $\det(\cdot)$ and $\rho_{\max}(\cdot)$ denote the trace, determinant, and maximum eigenvalue, respectively, of a square matrix; $\mathrm{tr}^q(\cdot) = (\mathrm{tr}(\cdot))^q$; $\|\cdot\|$ denotes the spectral norm, i.e., $\|\mathbf{A}\| = \rho_{\max}^{1/2}\left(\mathbf{A}'\mathbf{A}\right)$ for arbitrary $\mathbf{A}$; 1_h denotes a $h \times 1$ vector of ones; I_c denotes the $c \times c$ identity matrix; $\mathbf{0}_{c\times d}$ denotes a $c \times d$ matrix of zeros; and we use $n, T \to \infty$ to denote $n \to \infty$ and $T \to \infty$, jointly. Proofs are located in Appendix B.

2. ASSUMPTIONS AND FORECASTING MODELS

In this section we describe the class of panel data processes under consideration before defining the candidate forecasting models and least squares forecasts.

2.1. Assumptions

We observe a sample of $m \times 1$ vectors $\left\{y_{i,t}\right\}_{i,t=1}^{n,T}$ that are generated by an infinite order panel VAR process of the form

$$y_{i,t} = \sum_{r=1}^{p} \gamma_{i,r} (t-1)^{(r-1)} + x_{i,t},\ x_{i,t} = \sum_{s=1}^{\infty} \alpha_s' x_{i,t-s} + e_{i,t},\ t = \ldots, -1, 0, 1, \ldots;\ i = 1, 2, \ldots, \tag{1}$$

such that $x_{i,t}$ is the stochastic component of $y_{i,t}$ and $\sum_{r=1}^{p} \gamma_{i,r} (t-1)^{(r-1)}$ is the deterministic component. The vector autoregressive structure underlying the stochastic component is homogenous in the sense that $\{\alpha_s\}_{s=1}^{\infty}$ are the same for each $i = 1, 2, \ldots$. Under the assumptions stipulated below, the innovations to the VAR process $(e_{i,t})$ are independently distributed over time, but can exhibit weak cross sectional dependence. Meanwhile the deterministic component is a set of polynomial time trends interacting with cross section specific coefficients. It admits common forms of parameterized heterogeneity, such as cross sectional fixed effects $(p = 1)$ or heterogenous linear trends $(p = 2)$.

We impose the following assumptions on $\{\alpha_s\}_{s=1}^{\infty}$.

Assumption 1. *$det(I_m - \sum_{s=1}^{\infty} z^s \alpha_s) \neq 0$ for $|z| \leq 1$, $\sum_{s=1}^{\infty} s \|\alpha_s\| \leq C_\alpha < \infty$, and the diagonal elements of α_s are nonzero for infinitely many s.*

Under Assumption 1 the VAR process is invertible and can be expressed as a vector moving average process with moving average matrices $\{\theta_s\}_{s=0}^{\infty}$ recursively defined as $\theta_s := \sum_{r=0}^{s-1} \theta_r \alpha_{s-r}$ for $s \geq 1$ and $\theta_0 = \alpha_0 = I_m$. Moreover, the absolute summability condition placed on $\{\alpha_s\}_{s=1}^{\infty}$ implies that $\sum_{s=0}^{\infty} s \|\theta_s\|$ is also bounded (see Thm. 3.8.4 of Brillinger, 1981). This condition limits the dependence in the time series and ensures that LS estimator bias does not dominate the forecast variance in the asymptotic QFR (see the Remarks in Section 3.1 for further details).

We impose the following assumptions on the error process $e_{i,t}$.

Assumption 2.

(i) For each $i = 1, 2\ldots$, $\left\{e_{i,t}\right\}_{t=-\infty}^{\infty}$ is a sequence of independent random vectors with zero mean and variance $\Sigma_{i,i} > 0$ satisfying $\sup_{-\infty<t<\infty} \mathrm{E}\left\|e_{i,t}\right\|^q < C_q$ for $q = 1, 2, \ldots$

(ii) For each $i = 1, 2\ldots$ and $t = \ldots, -1, 0, 1, \ldots$, there exist positive constants b, c and d such that $\sup_{\underline{v}'\underline{v}=1} \mathrm{P}\left(a_1 < \underline{v}' e_{i,t} < a_2\right) \leq c(a_2 - a_1)^d$ for all a_1 and a_2 satisfying $0 < a_1 - a_2 \leq b$.

(iii) $\frac{1}{n} \sum_{i,j=1}^{n} \left\| \Sigma_{i,j} \right\| \leq C_\Sigma$ for all n, where $\Sigma_{i,j} := \mathrm{E}\left(e_{i,t} e_{j,t}'\right)$ for each $i, j = 1, 2, \ldots$

Under Assumption 2(i) all moments of the errors are bounded, a condition necessary to derive an asymptotic QFR expression under the relatively fast growth rate in model dimensionality permitted (see Assumption 3 below).[3] It also allows us to obtain asymptotic efficiency under the most general conditions on the relative

[3] Ing and Wei (2003) show that the bound can be relaxed at the expense of a slower rate of growth in the model dimension.

rates of expansion in n and T. Note that under (i), the sequence $\{e_{i,t}\}_{t=-\infty}^{\infty}$ is independently and heterogeneously distributed but with homogenous covariance $\Sigma_{i,i}$ for each $i = 1, 2, \ldots$. This means we impose time series homoskedasticity but permit cross sectional heteroskedasticity in $e_{i,t}$.[4] Assumption (ii) characterizes the distribution of the error vector as *uniformly Lipschitz*, and it ensures that the expectation of the smallest eigenvalue of the regressor covariance matrix is bounded above zero in sufficiently large samples, even as the lag order grows in the asymptotics (see, for example, Ing and Wei, 2003, 2005; Findley and Wei, 2002).

Assumptions 1, 2(i) and (ii) apply standard conditions imposed in the time series forecasting literature to each time series in the panel (see, for example, Ing and Wei, 2005). Assumption 2(iii) characterizes the cross-sectional dependence in the panel. Although $e_{i,t}$ is independently distributed over time under Assumption 2(i), it is not necessarily independently distributed between cross sections. Assumption 2(iii) limits the cross section correlation in $e_{i,t}$. Together with Assumption 1, Assumption 2(iii) ensures that the panel vectors $x_{i,t}$ obey *weak dependence* conditions of the form

$$\frac{1}{n}\sum_{i,j=1}^{n}\frac{1}{T}\sum_{s,t=1}^{T}\mathrm{E}\left(x'_{i,t}x_{j,s}\right) \leq \mathrm{tr}\left(\frac{1}{n}\sum_{i,j=1}^{n}\sum_{s,r=0}^{\infty}\theta'_s\Sigma_{i,j}\theta_r\right) \leq m\,\|C_\theta\|^2\,\|C_\Sigma\| < \infty \quad \mathbf{(2)}$$

(see Chudik, Pesaran, and Tosetti, 2011), where $\sum_{s=0}^{\infty}\|\theta_s\|^2 \leq C_\theta$ for some finite positive constant C_θ.

In the following subsection we will consider fitting models of finite lag order (denoted k) to the panel generated by (1). Because the panel process $y_{i,t}$ is a VAR(∞), these finite order fitted models are misspecified. For each $k = 1, 2, \ldots$ we define

$$s_{i,t}(k) := \sum_{s=1}^{\infty}(\alpha_s - \alpha_s(k))'\,x_{i,t-s+1},$$

where

$$\left(\{\alpha_s(k)\}_{s=1}^{k}\right) := \underset{a_s\in\mathbb{R}^{m\times m}}{\arg\min}\sum_{t=k}^{T-1}\sum_{i=1}^{n}\mathrm{E}\left(\left\|x_{i,t}-\sum_{s=1}^{k}a'_s x_{i,t-s}\right\|^2\right),\quad \alpha_s(k) := 0 \text{ for } s \geq k+1,$$

such that $s_{i,t}(k)$ represents the specification error of the VAR(k) model. We signify the parameters with k to indicate that the parameters that solve the minimization problem change with the number of lags. It will be convenient to decompose the vector $x_{i,t}$ as

$$x_{i,t} = \sum_{s=1}^{k}\alpha'_s(k)\,x_{i,t-s} + u_{i,t}(k), \quad \mathbf{(3)}$$

where $u_{i,t}(k) := e_{i,t} + s_{i,t-1}(k)$ for each $k = 1, 2, \ldots$, and it will also prove convenient to define $\boldsymbol{\alpha}(k) := \left[\alpha'_1(k) : \cdots : \alpha'_k(k)\right]'$.

[4] Homoskedasticity is commonly assumed in the time series forecasting literature. See, among others, Ing and Wei (2003; 2005), Ing (2003), and Schorfheide (2005).

2.2. Least Squares Forecasts

We produce out-of-sample forecasts using a set of finite order panel vector autoregressions (VARs) of the form

$$y_{i,t} = \sum\nolimits_{r=1}^{p} \beta_{i,r} (t-1)^{(r-1)} + \sum\nolimits_{s=1}^{k} \alpha_s' y_{i,t-s} + e_{i,t}, \quad \textbf{(4)}$$

where $k = 1, 2, \ldots, k_{n,T}$ indexes the different models up to some maximum lag order $k_{n,T}$. We begin by focussing on the OLS estimator of (4), which has several appealing features as a potential predictor. First, the estimator is consistent (for the population coefficients $\{\alpha_s(k)\}_{s=1}^{k}$) provided $T \to \infty$. Second, they are straightforward to compute, and are often employed in empirical panel forecasts (see the "Forecasting Applications" section of Baltagi, 2008, and the references cited therein). Later, in Section 3.5, we will consider bias-corrected LS estimators of (4).

The set of OLS estimates are defined as follows. For all integers k and l satisfying $1 \le k \le l \le k_{n,T}$, we have

$$\left(\hat{\alpha}(k,l), \left\{\hat{\beta}_i(k,l)\right\}_{i=1}^{n}\right) := \underset{\mathbf{a}\in\mathbb{R}^{mk\times m}, \mathbf{b}_i\in\mathbb{R}^{p\times m}}{\arg\min} \sum_{t=l}^{T-1} \sum_{i=1}^{n} \left\| y_{i,t+1} - \mathbf{a}' Y_{i,t}(k) - \mathbf{b}_i' \tau_{t-l+1} \right\|^2, \quad \textbf{(5)}$$

where $Y_{i,t}(k) := \left(y_{i,t}', \ldots, y_{i,t-k+1}'\right)'$ is a $mk \times 1$ vector and $\tau_t := \left(1, t-1, \ldots, (t-1)^{p-1}\right)'$ is a $p \times 1$ vector of polynomial time trends. Note that (5) corresponds to an equation-by-equation OLS estimation of the VAR(k).

Remark. Note that l in (5) dictates the beginning of the time series sample used in estimation of the VAR(k). Specifically, the sample of dependent variables used in estimation of the model is given by $\left\{y_{i,t}\right\}_{i=1,t=l+1}^{n,T}$. In what follows we focus on two specific choices of l. For the purpose of producing a forecast, we will set $l = k$, so that the VAR(k) is fitted to the full available sample of data (see (6) below). That is, the sample of dependent variables used in estimation of the model varies with k, and is given by $\left\{y_{i,t}\right\}_{i=1,t=k+1}^{n,T}$. However, for the purpose of estimating the QFR of these forecasts, we fit each VAR(k) to a common sample by setting $l = k_{n,T}$ (see Section 3.2 below). That is, $\left\{y_{i,t}\right\}_{i=1,t=k_{n,T}+1}^{n,T}$ comprises the sample of dependent variables used to estimate all models. This common sample is critical for establishing the asymptotic efficiency of the model selection rule (see Remark (i) in Section 3.3 below for further details).[5]

[5] We use the full sample of data to construct forecasts because estimating parameters from the common sample would lower the precision of the point estimates and inflate the quadratic loss of the forecasts. This restriction would be particularly costly for the precision of the fitted heterogenous parameters, each of which are identified only as T grows large.

For each $i = 1, \ldots, n$, the OLS out-of-sample forecast from the fitted VAR(k) is

$$\hat{y}_{i,T+1}(k) = \hat{\alpha}(k,k)' Y_{i,T}(k) + \hat{\beta}_i(k,k)' \tau_{T-k+1}, \tag{6}$$

and the associated *quadratic forecast loss* (QFL) is defined as $\mathrm{tr}(\Phi \mathcal{L}(k))$, where

$$\mathcal{L}(k) := \frac{1}{n} \sum_{i=1}^{n} \left(\hat{y}_{i,T+1}(k) - y_{i,T+1}\right)\left(\hat{y}_{i,T+1}(k) - y_{i,T+1}\right)',$$

and Φ is a positive semidefinite (PSD) $m \times m$ weighting matrix satisfying $\|\Phi\| = 1$. The *quadratic forecast risk* (QFR) of the model is then $\mathrm{E}(\mathrm{tr}(\Phi \mathcal{L}(k)))$. Note that the OLS model parameters used in the forecast (6) are estimated using the full available sample, since we have set $l = k$ in (5) when constructing the forecast.

In general Φ is chosen by the practitioner according to which elements (or combinations of elements) of $y_{i,T+1}$ they are interested in forecasting (also see Schorfheide, 2005, and Greenaway-McGrevy, 2013; 2015). This approach to evaluating the quadratic loss of the forecast is rather flexible. Because Φ is PSD, there exists $\imath$ such that $\Phi = \imath\imath'$, which allows us to study the quadratic loss of linear combinations of the form $\imath' \hat{y}_{i,T+1}(k)$. For example, if we are interested in forecasting the first element of $y_{i,T+1}$, then $\imath = (1, 0, \ldots, 0)'$. If we are interested in forecasting the difference between the first and second elements of $y_{i,T+1}$, then $\imath = \left(\sqrt{1/2}, -\sqrt{1/2}, 0, \ldots, 0\right)'$. The condition $\|\Phi\| = 1$ is a straightforward normalization that is made for expositional clarity.

Exogenous regressors can also be accommodated in the analysis by expanding the vector $y_{i,t}$ to include the exogenous variables, provided that the exogenous variables obey the conditions stipulated under Assumptions 1 and 2. In particular, the exogenous regressors must be weakly dependent in both the time series and cross sectional dimensions of the panel. By definition, exogenous regressors would impose restrictions on the parameter space of the data generating process, such as block zero restrictions on the off-diagonal elements of $\{\alpha_s\}_{s=1}^{\infty}$. Because such restrictions can easily be accommodated under Assumption 1, the main results of the article will continue to apply.

The fitted parameters $\hat{\alpha}(k,k)$ and $\left\{\hat{\beta}_i(k,k)\right\}_{i=1}^{n}$ in (6) are expressed as functions of the lag order k since our focus is the effect of model specification on model fit. Because $\hat{\beta}_i(k,k)$ are specific for each i, they converge at a $\sqrt{T}$ rate (for a given k). By contrast, $\hat{\alpha}(k,k)$ is homogenous and is estimated by pooling across i and t, and hence convergence would occur at a faster $\sqrt{nT}$ rate when k is fixed (Hahn and Kuersteiner, 2002; Alvarez and Arellano, 2003). The $O\left(T^{-1}\right)$ variance of the estimated cross section specific parameters therefore dominates the $O\left((nT)^{-1}\right)$ variance of the estimated homogenous coefficients in the QFR. Although the number of these heterogenous parameters is not dependent on the lag order, k does have a smaller $O\left(kT^{-2}\right)$ effect on the variance of these fitted coefficients since the number of observations used in estimation is $T - k$ (see the

Remarks in Section 3.1 for more details). Hence, in order to discern the relationship between k and forecast risk, these smaller $O\left(kT^{-2}\right)$ terms in the QFR will be taken into account.

We only consider forecasting one period ahead in (6). While we do not consider multistep forecasting in this article, the theory provides the basis for the development of multistep forecasting selection that will be pursued in future work. Also see Greenaway-McGrevy (2013) for a discussion of the OLS bias in correctly specified multistep panel data VARs.

The variables being forecast in (6), namely $\left\{y_{i,T+1}\right\}_{i=1}^{n}$, are generated from the same underlying process as the data used to estimate model parameters, namely $\left\{y_{i,t}\right\}_{i,t=1}^{n,T}$. This corresponds to *same-sample* realization, and it contrasts against *independent sample* realization, wherein the process to be forecast shares the same stochastic structure but is independent of the data used to fit the model.[6] Independent realization is mathematically convenient because it permits us to ignore any dependence between the estimated model parameters and the predictor variables when solving for the QFR of the forecast (see, among others, Shibata, 1980; Bhansali, 1996, 1997; Schorfheide, 2005). However, the independent realization assumption is untenable in many empirical applications, and while it is an innocuous assumption in many time series applications, it leads to an altogether different expression for the asymptotic QFR compared to same-sample realization in panel data settings (Greenaway-McGrevy, 2015).

We restrict the maximum permissible lag order $k_{n,T}$ as follows.

Assumption 3. *As $n \to \infty$ and $T \to \infty$, the set of lags orders is $k \in \left\{1, 2, \ldots, k_{n,T}\right\}$, where $k_{n,T}$ is a sequence of positive integers satisfying either*

$$C_1 \leq \lim_{n,T \longrightarrow \infty} \left(\frac{k_{n,T}^{1+\epsilon}}{\sqrt{n}} + \frac{k_{n,T}^{1+\epsilon}}{T} \right) \leq C_2 \tag{7}$$

or

$$C_1 \leq \lim_{n,T \longrightarrow \infty} \left(\frac{k_{n,T}^{1+\epsilon}}{\sqrt{T}} \right) \leq C_2 \tag{8}$$

for some $\epsilon > 0$ and positive constants C_1 and C_2.

The restriction imposed on $k_{n,T}$ under either (7) or (8) ensures that the dependence between the estimated model parameters (i.e., $\hat{\alpha}(k,k)$ and $\left\{\hat{\beta}_i(k,k)\right\}_{i=1}^{n}$) and $Y_{i,T}(k)$ in (6) is asymptotically negligible in the same-sample framework, albeit under very different mechanisms. If (7) holds, then it follows that $k_{n,T} n^{-1/2} = o(1)$, and the dependence is shown to be negligible by exploiting the weak cross

[6] Suppose we have scalars $z_{1,t} = \sum_{s=0}^{\infty} \eta_s w_{1,t-s}$ and $z_{2,t} = \sum_{s=0}^{\infty} \eta_s w_{2,t-s}$ for some $\{\eta_s\}_{s=0}^{\infty}$, where $w_{1,t}$ and $w_{2,t}$ are white noise with identical variances but are statistically independent of each other. Fitting an autoregressive model to $\left\{z_{1,t}\right\}_{t=1}^{T}$ and using the fitted coefficients to forecast $z_{2,T+1}$ is an example of independent realization. Using those same parameters to forecast $z_{1,T+1}$ is an example of same-sample realization.

sectional dependence in the panel. This condition is less restrictive than (8) in "large n" panels ($n >> T$). If instead (8) holds, then it follows that $k_{n,T}T^{-1/2} = o(1)$, and the dependence is shown to negligible by exploiting the weak time series dependence in the panel, as in Ing and Wei (2005). This condition is less restrictive than (7) in "large T" panels ($T >> n$). For comparison, in the time series context the maximum permissible rate of expansion in the lag order is restricted to be $o\left(T^{1/2}\right)$ (Shibata, 1980; Ing and Wei, 2005), meaning that we can permit much larger models for panels with n larger than T.

3. ASYMPTOTIC THEORY

In this section we derive an asymptotic expression for the QFR of the LS forecasts defined in (6) above. Based on this expression, we construct an asymptotically unbiased estimator of the QFR that provides the basis for the new model selection method. We then provide conditions under which the model selection method is asymptotically efficient. Finally, we consider an extension to bias-corrected least squares forecasts.

3.1. Asymptotic Quadratic Forecast Risk

Our first step towards developing forecast model selection criteria is to solve for an analytic expression for the QFR. Before proceeding, we introduce the following notation. Let

$$\Gamma_{i,j}(k) := \mathrm{E}\left(X_{i,t}(k)\, X'_{j,t}(k)\right) \textit{ for each } i, j = 1, 2, \ldots,$$

where $X_{i,t}(k) := \left(x'_{i,t}, \ldots, x'_{i,t-k+1}\right)'$, and[7]

$$\Gamma(k) := \frac{1}{n}\sum_{i=1}^{n} \Gamma_{i,i}(k), \quad \Sigma := \frac{1}{n}\sum_{i=1}^{n} \Sigma_{i,i}.$$

Theorem 3.1 below characterizes $\operatorname{tr}\left(\Phi L_{n,T}(k)\right)$ as the (second order) asymptotic QFR, where

$$L_{n,T}(k) := \Lambda(k) + \Sigma\left[\frac{kp^2}{T(T-k)} + \sum_{l=1}^{p} \frac{2l-1}{T-k}\left(\frac{\prod_{r=0}^{l-1}(T-k+r)}{\prod_{r=0}^{l-1}(T-k-r)} - 1\right)\right] + \Pi(k)\frac{k}{n(T-k)} + \zeta(k)\frac{kp^2}{(T-k)^2}, \tag{9}$$

for

$$\Lambda(k) := \sum_{s,r=1}^{\infty} (\alpha_s(k) - \alpha_s)'\left(\frac{1}{n}\sum_{i=1}^{n} \mathrm{E}\left(x_{i,t-s}x'_{i,t-r}\right)\right)(\alpha_r(k) - \alpha_r),$$

$$\Pi(k) := \frac{1}{nk}\sum_{i,j=1}^{n} \Sigma_{i,j} \cdot \operatorname{tr}\left(\Gamma_{i,j}(k)\,\Gamma^{-1}(k)\right), \tag{10}$$

[7] Note that due to potential cross sectional heterogeneity in $e_{i,t}$, Σ and $\Gamma(k)$ can depend on n. In the interests of brevity this dependence is not made explicit in the notation.

and

$$\zeta(k) := \tfrac{1}{k}\Sigma\left(1_k \otimes \left(I_m - \sum_{s=1}^{\infty}\alpha_s'\right)^{-1}\right)'\Gamma^{-1}(k)\left(1_k \otimes \left(I_m - \sum_{s=1}^{\infty}\alpha_s'\right)^{-1}\right)\Sigma. \tag{11}$$

Note that $L_{n,T}(k)$ is a sum of positive definite matrices, and hence $L_{n,T}(k) > 0$ (in the matrix sense). Although $L_{n,T}(k)$ appears to be rather complicated, each term comprising the sum has a natural interpretation, as discussed in the following remarks.

Remarks. (i) The term $\Lambda(k)$ represents the effect of specification error in the VAR(k) model on the QFR (it is sometimes referred to as *goodness of fit*). Mappings from $\Lambda(k)$ to the real line are a measure of distance between the regression model (4) and the data generating process.[8] Note that while $\mathrm{tr}(\Phi\Lambda(k)) > 0$ for all $k \geq 1$ under Assumption 1, the term is decreasing in k, corresponding to the fact that the specification error of the model decreases as the fitted lag order increases.[9]

(ii) The remaining terms in the expression for $L_{n,T}(k)$ are the *model complexity* terms that are increasing k, and these terms reflect estimation error in the fitted model parameters. The term

$$\Sigma\left[\frac{kp^2}{T(T-k)} + \sum_{l=1}^{p}\frac{2l-1}{T-k}\left(\frac{\prod_{r=0}^{l-1}(T-k+r)}{\prod_{r=0}^{l-1}(T-k-r)} - 1\right)\right]$$

arises from the second order effect of variance in the fitted heterogenous coefficients $\left\{\hat{\beta}_i(k,k)\right\}_{i=1}^{n}$ on the QFR. The first order effect is $\frac{p^2}{T}\Sigma$, which is independent of the lag order k, and hence the term will be subtracted from $\mathcal{L}(k)$ before deriving the asymptotic QFR (see Theorem 3.1 below). However, the variance of these fitted coefficients does have a second order impact on the QFR because the effective number of time series observations with which to estimate these parameters (i.e., $T-k$) is decreasing in k.[10]

(iii) The term $\frac{k}{n(T-k)}\Pi(k)$ captures the effect of variance in the fitted homogenous parameters $\hat{\boldsymbol{\alpha}}(k,k)$ on the QFR. The term is increasing in k because $\Pi(k) = \frac{1}{nk}\sum_{i,j=1}^{n}\Sigma_{i,j} \times \mathrm{tr}\left(\Gamma_{i,j}(k)\Gamma^{-1}(k)\right) = O(1)$ under weak cross sectional dependence (see Assumption 2(iii)). Note that any weak correlation between cross sectional units will manifest itself in this term.

---

[8] $\Lambda(k)$ corresponds to the term $\|\boldsymbol{\alpha}(k) - \boldsymbol{\alpha}\|_{\Gamma}^2$ used by Ing and Wei (2003, 2005) and Shibata (1980) for the time series autoregressive case.

[9] Ing and Wei (2003) show that when $m = 1$, $\Lambda(k) \leq C\sum_{s=k+1}^{\infty}\theta_s^2$. Thus if $\sum_{s=1}^{\infty}\sqrt{s}\,\|\theta_s\| < \infty$, for example, then $\Lambda(k) \leq Ck^{-1}$.

[10] Note that the QFR is quadratic in the order of the polynomial trend (i.e., $p-1$). Corollary 1 of Ing (2003) demonstrates a similar result in the times series case.

(iv) The final term in (9), namely $\frac{kp^2}{(T-k)^2}\zeta(k)$, captures the effect of the $O\left(T^{-1}\right)$ bias (or *Nickell bias*) in LS estimator on the QFR. For example, we can express

$$\hat{\boldsymbol{\alpha}}(k,k)-\boldsymbol{\alpha}(k)=\tfrac{p}{T-k}\Gamma^{-1}(k)\boldsymbol{\xi}(k)+d_{n,T}(k),\ \boldsymbol{\xi}(k):=\left(1_k\otimes\left(I_m-\textstyle\sum_{s=1}^{\infty}\alpha_s'\right)^{-1}\right)\Sigma,$$

where $d_{n,T}(k)$ denotes an asymptotically negligible term that can be ignored for our purposes.[11] In correctly specified models, $\boldsymbol{\xi}(k)$ simplifies to the approximations derived elsewhere (e.g., Hahn and Kuersteiner, 2002; Greenaway-McGrevy, 2013). The $\zeta(k)$ term is then a quadratic transformation of the Nickell bias, since $\zeta(k)=\frac{1}{k}\boldsymbol{\xi}(k)'\Gamma^{-1}(k)\boldsymbol{\xi}(k)$. The Nickell bias induces an $O\left(k(T-k)^{-2}\right)$ term in the QFR, since $k\times\zeta(k)$ is nondecreasing in k. To see why this is the case, note that we can decompose $\Gamma^{-1}(k)=\mathbf{G}(k)\mathbf{S}^{-1}(k)\mathbf{G}'(k)$, for a $mk\times mk$ block lower triangular matrix $\mathbf{G}(k)$ with I_m along the principal diagonal and matrices $-\alpha_{h-1}(k-l)$ in the (h,l) block $(h>l)$; and block diagonal $\mathbf{S}(k)$ with $\Sigma(l):=\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}u_{i,t}(k-l)u_{i,t}'(k-l)\right)$ in the (l,l) diagonal block, where $u_{i,t}(0):=x_{i,t}$ (cf. (2.18) in Bhansali, 1996). Then we have $k\zeta(k)=\sum_{h=1}^{k}\boldsymbol{\Upsilon}_h$, where

$$\boldsymbol{\Upsilon}_h:=\Sigma\left(I_m-\sum_{s=1}^{\infty}\alpha_s\right)^{-1}\left(I_m-\sum_{s=1}^{h}\alpha_s(h)\right)\Sigma^{-1}(k)\left(I_m-\sum_{s=1}^{h}\alpha_s'(h)\right)\left(I_m-\sum_{s=1}^{\infty}\alpha_s'\right)^{-1}\Sigma$$

satisfies $\boldsymbol{\Upsilon}_h\geq 0$ (in the matrix sense). The quadratic bias term $\zeta(k)$ is also dependent on the summability condition imposed on $\{\alpha_s\}_{s=1}^{\infty}$ under Assumption 1. For highly dependent processes that violate this summability condition, we cannot ensure that the quadratic bias term is $O\left(k(T-k)^{-2}\right)$. Finally, note that in the time series context such bias terms are not included in the asymptotic expression of the QFR, despite the fact that it is well-known that the ordinary least squares (OLS) estimator of autoregressions exhibits $O\left(T^{-1}\right)$ bias. This is because the effect of the bias on the relevant approximation of the QFR is of smaller order than the effect of estimator variance.

The following result characterizes $L_{n,T}(k)$ as the asymptotic QFR.

[11] Specifically, $\left\|\mathrm{E}\left(d_{n,T}(k)\right)\right\|^2$ is proportional to $k(T-k)^{-2}\|\Lambda(k)\|$, meaning the effect of $d_{n,T}(k)$ on quadratic forecast loss is dominated by $\Lambda(k)$ as $T\to\infty$, and hence the stochastic term $d_{n,T}(k)$ need not be considered when deriving the asymptotic QFR expression.

THEOREM 3.1. *Let $y_{i,t}$ be generated according to (1). Then under Assumptions 1, 2, and 3,*

$$\lim_{n,T\to\infty} \max_{1\le k\le k_{n,T}} \left| \frac{\operatorname{tr}\left(\Phi\left[\mathrm{E}\left(\mathcal{L}(k)\right) - \left(1+\frac{p^2}{T}\right)\Sigma\right]\right)}{\operatorname{tr}\left(\Phi L_{n,T}(k)\right)} - 1 \right| = 0 \tag{12}$$

for all positive semidefinite $m\times m$ matrices Φ satisfying $\|\Phi\| = 1$.

Theorem 3.1 states that the second order QFR of the LS forecast can be uniformly (in $k \le k_{n,T}$) approximated by $L_{n,T}(k)$. It is second order in the sense that we subtract $\operatorname{tr}\left(\Phi\Sigma\left(1+T^{-1}p^2\right)\right)$ from the QFR before taking the limit. Note that Σ is the average (across i) covariance of the innovation vector $e_{i,T+1}$, while $\Sigma p^2 T^{-1}$ is the $O\left(T^{-1}\right)$ approximation of the average (across i) covariance of the estimated cross section specific coefficients $\{\hat{\alpha}_i(k,k)\}_{i=1}^n$. Both terms are independent of the lag order k, and hence we must consider the smaller second order terms contained in $L_{n,T}(k)$ for the purpose of model selection.

The proof of Theorem 3.1 is located in the Appendix. In order to solve for $\mathrm{E}(\mathcal{L}(k))$, we require that the minimum eigenvalue of the sample covariance matrix of the regressors is bounded above zero in expectation, even as $k_{n,T}$ grows large. Findley and Wei (2002) and Ing and Wei (2003) show that the Lipschitz assumption (Assumption 2(ii)) ensures that the expectation is bounded away from zero in sufficiently large samples.

The relative rates of expansion in n and T are not restricted, but they will affect the relative magnitudes of the additive model complexity terms in the asymptotic QFR. For example, if n grows sufficiently fast, such that $T = o(n)$, then the quadratic bias and the second-order variance of the cross section specific parameters $\left\{\hat{\beta}_i(k,k)\right\}_{i=1}^n$ comprise the dominant model complexity terms. Conversely, if $n = o(T)$, the variance of the homogenous VAR coefficients $\hat{\alpha}(k,k)$ comprise the dominant model complexity term. Under either condition the asymptotic QFR can be simplified considerably (see Section 3.4 for further details). If n and T grow at the same rate (e.g., $n/T \to C$, $0 < C < \infty$), then all terms are manifest in the asymptotic QFR.

Having established $\operatorname{tr}\left(\Phi L_{n,T}(k)\right)$ as the asymptotic QFR of the LS forecast, we define

$$k^*_{n,T} := \arg\min_{1\le k\le k_{n,T}} \operatorname{tr}\left(\Phi\, L_{n,T}(k)\right).$$

Under Assumption 1, using the same arguments as Shibata (1980, p. 154) we can establish that $k^*_{n,T} \to \infty$ as $n\to\infty$ and $T\to\infty$. As n and T grow large, the sequence of $k^*_{n,T}$ provides the basis for evaluating model selection methods. Asymptotically efficient model selection methods choose a sequence of QFRs that converges to the sequence $\operatorname{tr}\left(\Phi\, L_{n,T}\left(k^*_{n,T}\right)\right)$ as the sample size increases. Note that $k^*_{n,T}$ is implicitly dependent on Φ and the relative rates of expansion in n and T.

3.2. Quadratic Forecast Risk Estimation and Model Selection

Estimators of quadratic forecast risk are often used as the basis for model selection methods (e.g., Akaike, 1970; Shibata, 1980; Ing and Wei, 2005; Schorfheide, 2005). The underlying intuition is to estimate the quadratic forecast risk for each model within the candidate set, and then select the model corresponding to the smallest estimated risk.

In order to estimate the QFR of each model, we fit each VAR(k) to a common sample by setting $l = k_{n,T}$ in (5) above for each $k = 1, 2, \ldots, k_{n,T}$. Thus, the sample of dependent variables used in estimation of each model is $\{y_{i,t}\}_{i=1,t=k_{n,T}+1}^{n,T}$. This common sample is necessary for the asymptotic efficiency of the model selection criterion (see the discussion after Theorem 3.2 below).

The model selection criteria proposed by Akaike (1970), Shibata (1980) and Mallows (1973) are based on penalizing the in-sample quadratic loss of the fitted model. We consider a similar approach. First, for $i = 1, \ldots, n$ and $t = l, \ldots, T$, we define

$$\hat{u}_{i,t+1}(k,l) := \hat{y}_{i,t+1}(k,l) - y_{i,t+1}, \quad \hat{y}_{i,t+1}(k,l) := \hat{\alpha}(k,l)' Y_{i,t}(k) + \hat{\beta}_i(k,l)' \tau_{t-l+1},$$

for all k and l satisfying $1 \le k \le l \le k_{n,T}$. We then let $\mathrm{tr}\left(\Phi \hat{R}(k)\right)$ denote the in-sample quadratic loss of the VAR(k) fitted to $\{y_{i,t}\}_{i=1,t=k_{n,T}+1}^{n,T}$, where

$$\hat{R}(k) := \frac{1}{n(T-k_{n,T})} \sum_{t=k_{n,T}}^{T-1} \sum_{i=1}^{n} \hat{u}_{i,t+1}\left(k, k_{n,T}\right) \hat{u}'_{i,t+1}\left(k, k_{n,T}\right). \tag{13}$$

Our suggested QFR estimator is then

$$\hat{L}_{n,T}(k) := \hat{R}(k) + \left[\hat{R}(k) \frac{T-k_{n,T}}{T-k_{n,T}-p}\right] \left(\frac{p}{T-k_{n,T}} + \sum_{l=1}^{p} \frac{2l-1}{T-k} \frac{\prod_{r=0}^{l-1}(T-k+r)}{\prod_{r=0}^{l-1}(T-k-r)} \right) + \hat{\Pi}(k) \frac{k}{n} \left[\frac{1}{T-k_{n,T}} + \frac{1}{T-k} \right] + \hat{\zeta}(k) \left[\frac{kp^2}{(T-k_{n,T})^2} + \frac{kp^2}{(T-k)^2} \right], \tag{14}$$

where $\hat{\Pi}(k)$ and $\hat{\zeta}(k)$ denote estimates of $\Pi(k)$ and $\zeta(k)$ (specified in (17) and (18) below), and where $\hat{R}(k) \frac{T-k_{n,T}}{T-k_{n,T}-p}$ is used as an estimate of Σ. Following the terminology set forth by Akaike (1970), we refer to this estimator as *panel final prediction error* (panel FPE).[12]

Our model selection rule is to choose k to minimize the estimated QFR, i.e.,

$$\hat{k} := \arg \min_{1 \le k \le k_{n,T}} \mathrm{tr}\left(\Phi \hat{L}_{n,T}(k)\right). \tag{15}$$

This criterion is similar to model selection by minimization of estimated out-of-sample risk in the time series literature (Akaike, 1970; Shibata, 1980). However, in contrast to these criteria, minimizing the panel FPE imposes an additional

[12] The design of $\hat{L}_{n,T}(k)$ is based on comparing $L_{n,T}(k)$ to the asymptotic expression for $\hat{R}(k)$ provided in Lemma B.5 of the Appendix.

penalty on overfitting through the final term in (14), which accounts for the effect of LS estimator bias on forecast loss. The second term in (14) ensures that the selection criterion (15) also imposes an additional penalty on model dimension to account for the effect of variability in the fitted heterogenous parameters (such as fixed effects) on forecast loss. The remainder of this section focuses on how to construct $\hat{\Pi}(k)$ and $\hat{\zeta}(k)$.

First we consider an estimate of $\Pi(k)$. A key feature of our framework is that the error vectors $e_{i,t}$ are permitted to be weakly dependent in the cross sectional dimension (but note that they are uncorrelated across time). We therefore treat each time period in the panel as a cluster when estimating $\Pi(k)$.[13] We require some preliminary notation before introducing the clustering estimator. Let

$$\mathbf{Y}_i^{(k,l)} := \left[Y_{i,l}(k) : \cdots : Y_{i,T-1}(k)\right]', \; \hat{\mathbf{Q}}_{(k,l)} := \frac{1}{n(T-l)} \sum_{i=1}^{n} \mathbf{Y}_i^{(k,l)\prime} \mathbf{M}_{T-l} \mathbf{Y}_i^{(k,l)}, \tag{16}$$

where $\mathbf{M}_N := I_N - \varsigma_N \left(\varsigma_N' \varsigma_N\right)^{-1} \varsigma_N'$ for all $N > p$ and $N \times p$ matrices $\varsigma_N := [\tau_1 : \cdots : \tau_N]'$. Thus $\mathbf{Y}_i^{(k,l)}$ is a $(T-l) \times km$ matrix of regressors for the ith cross sectional unit in a VAR(k) fitted to $\left\{y_{i,t}\right\}_{i=1,t=l+1}^{n,T}$, and $\mathbf{M}_{T-l}\mathbf{Y}_i^{(k,l)}$ is the matrix of regressors after partialing out the polynomial trend. The estimator of $\Pi(k)$ is then

$$\hat{\Pi}(k) := \frac{1}{n(T-k_{n,T})k} \sum_{i,j=1}^{n} \sum_{t=k_{n,T}}^{T-1} \left[\hat{u}_{i,t+1}(k,k_{n,T})\,\hat{u}_{j,t+1}'(k,k_{n,T})\,\mathrm{tr}\left(\ddot{Y}_{i,t}(k,k_{n,T})\,\ddot{Y}_{j,t}'(k,k_{n,T})\hat{\mathbf{Q}}_{(k,k_{n,T})}^{-1}\right)\right], \tag{17}$$

where $\ddot{Y}_{i,t}'(k,l)$ denotes the $(t-l+1)$th row of $\mathbf{M}_{T-l}\mathbf{Y}_i^{(k,l)}$ for $t = l, \ldots, T-1$.

Next, we require an estimate of the quadratic bias term $\zeta(k)$. Kiviet (1995) and Hahn and Kuersteiner (2002, 2011) use the fitted model parameters to provide a first order approximation of the $O\left(T^{-1}\right)$ bias of the LS estimator. We use a similar approach to estimate the quadratic bias function $\zeta(k)$ as follows:

$$\hat{\zeta}(k) := \frac{1}{k}\hat{\boldsymbol{\xi}}_{(k,k_{n,T})}' \hat{\mathbf{Q}}_{(k,k_{n,T})}^{-1} \hat{\boldsymbol{\xi}}_{(k,k_{n,T})}, \tag{18}$$

where

$$\hat{\boldsymbol{\xi}}_{(k,l)} := \left(I_{mk} - \hat{\mathbf{A}}_{(k,l)}'\right)^{-1} \mathbf{J}_{mk,m}\mathbf{J}_{mk,m}' \left(\hat{\mathbf{Q}}_{(k,l)} - \hat{\mathbf{A}}_{(k,l)}'\hat{\mathbf{Q}}_{(k,l)}\hat{\mathbf{A}}_{(k,l)}\right)\mathbf{J}_{mk,m} \tag{19}$$

for all integers k and l satisfying $1 \le k \le l \le k_{n,T}$. Here $\mathbf{J}_{d,c} := \left[I_c : \mathbf{0}_{c\times(d-c)}\right]'$ for $c \le d$, and the $mk \times mk$ matrix $\hat{\mathbf{A}}_{(k,l)}$ denotes the companion form of $\hat{\alpha}(k,l)$ (i.e., $\hat{\mathbf{A}}_{(k,l)} := [\hat{\alpha}(k,l) : \mathbf{L}_{mk}]$, where $\mathbf{L}_{mk} := \left[I_{m(k-1)} : \mathbf{0}\right]'$ is a $km \times (k-1)m$ matrix).

[13] Hansen (2007) shows that the covariance matrix of the OLS estimator can be consistently estimated by clustering within the dependent dimension of the panel, provided that both n and T grow large in the asymptotics. In his framework there is time series dependence, but the cross sectional units are independently distributed, so that consistency is achieved by clustering within each cross section. Here, we cluster within each time period since we permit cross sectional dependence.

In the Appendix we show that $\hat{L}_{n,T}(k)$ is an asymptotically unbiased estimator of the second order asymptotic QFR uniformly in $k \leq k_{n,T}$ (Theorem B.4). This result provides the theoretical justification for the model selection criterion given in (15).

3.3. Asymptotic Efficiency

In this subsection we provide conditions under which model selection by panel FPE-minimization (15) is asymptotically efficient in the sense defined by Shibata (1980). Following Ing and Wei (2005), we require the following condition.

Assumption 4. *For every $\xi > 0$ and positive semidefinite matrix Φ satisfying $\|\Phi\| = 1$, there exists π satisfying $0 < \pi < 1$ such that for all $k \in \left\{k : \left|k - k^*_{n,T}\right| > \left(k^*_{n,T}\right)^{\pi}\right\}$,*

$$k^{*\xi}_{n,T} \frac{\operatorname{tr}\left(\Phi L_{n,T}(k)\right) - \operatorname{tr}\left(\Phi L_{n,T}\left(k^*_{n,T}\right)\right)}{\delta^2_{n,T}\left|k - k^*_{n,T}\right|} \geq C,$$

where $\delta_{n,T} := \min\left(n^{\frac{1}{2}}\left(T - k_{n,T}\right)^{\frac{1}{2}}, T - k_{n,T}\right)$ and C is a positive constant.

This condition is required under same-sample realization, and it ensures that the probability that panel FPE-minimization selects a model outside the neighborhood of $k^*_{n,T}$ is arbitrarily small as the sample size grows large. It imposes a basin-like shape on the quadratic loss function, and as discussed in detail in Ing and Wei (2005), many stationary time series obey this condition, including invertible autoregressive moving average (ARMA) processes.

We then have the following.

THEOREM 3.2. *Let $y_{i,t}$ be generated according to (1), and assume that there exists $\upsilon \in (0,1)$ such that either $n^{\upsilon-1}T = o(1)$ or $T^{\upsilon-1}n = o(1)$ as $n \to \infty$ and $T \to \infty$. Then under Assumptions 1, 2, 3, and 4,*

$$\lim_{n,T\to\infty} \max_{1\leq k\leq k_{n,T}} \frac{\operatorname{tr}\left(\Phi\left[\mathrm{E}\left(\mathcal{L}\left(\hat{k}\right)\right) - \left(1 + \frac{p^2}{T}\right)\Sigma\right]\right)}{\operatorname{tr}\left(\Phi L_{n,T}\left(k^*_{n,T}\right)\right)} = 1 \tag{20}$$

for all positive semidefinite $m \times m$ matrices Φ satisfying $\|\Phi\| = 1$.

Theorem 3.2 provides conditions under which the model selection rule given in (15) is *asymptotically efficient*; i.e., as the sample size grows large, the selection rule minimizes the quadratic forecast risk within the candidate set of models. Note that the expectation $\mathrm{E}\left(\mathcal{L}\left(\hat{k}\right)\right)$ in (20) is taken over random $\hat{k}$.

Remarks. (i) The asymptotic efficiency of the model selection criterion (15) requires that the QFR estimators are based on a common sample.

Intuitively, by basing $\left\{\hat{L}_{n,T}(k)\right\}_{k=1}^{k_{n,T}}$ on a common sample, we eliminate some of the random variation between $\hat{L}_{n,T}(k)$ and $\hat{L}_{n,T}\left(k^*_{n,T}\right)$ for each $k \neq k^*_{n,T}$, making it easier to identify the effect of model dimensionality of forecast risk. More specifically, the common sample is necessary for $\mathrm{E}\left(\left\|\hat{L}_{n,T}(k) - \hat{L}_{n,T}\left(k^*_{n,T}\right)\right\|^q\right)$ to approach zero faster than $\left\|L_{n,T}(k)\right\|^q$ for all $k \leq k_{n,T}$ and any $q \geq 1$ as $n \to \infty$ and $T \to \infty$. See the proof of Theorem 3.2 in Appendix B for details. Han, Phillips, and Sul (2017) also discuss the need for a common sample when building a consistent model selection rule for panel data models with incidental parameters.

(ii) The restriction on the relative rates of growth in n and T under Theorem 3.2 serves a similar purpose. Specifically, in the proof to Theorem 3.2 we show that for all $q \geq 1$,

$$\mathrm{E}\left\|\hat{L}_{n,T}(k) - \hat{L}_{n,T}\left(k^*_{n,T}\right)\right\|^q = O\left(\frac{k^q}{n^{q/2}T^{3q/2}}\right) + O\left(\frac{\left(k^*_{n,T}\right)^q}{n^{q/2}T^{3q/2}}\right) + o\left(\left\|L_{n,T}(k)\right\|^q\right),$$

where recall that $L_{n,T}(k) = O\left((nT)^{-1}k\right) + O\left(T^{-2}k\right) + \Lambda(k)$.[14] Thus, if either $n^{\upsilon-1}T = o(1)$ or $T^{\upsilon-1}n = o(1)$ for some $\upsilon \in (0,1)$, the first two terms on the right hand side of the above equation are also $o\left(\left\|L_{n,T}(k)\right\|^q\right)$, and $\mathrm{E}\left\|\hat{L}_{n,T}(k) - \hat{L}_{n,T}\left(k^*_{n,T}\right)\right\|^q$ approaches zero faster than $\left\|L_{n,T}(k)\right\|^q$. Note the restriction precludes n and T growing at the same rate, but it can be satisfied if n grows either marginally slower or marginally faster than T.

(iii) As in time series treatments (e.g., Ing and Wei, 2005), the asymptotic efficiency result requires that the VAR(∞) does not degenerate to a finite-order VAR. In such cases, Lee and Phillips (2015) show that a Schwarz criterion modified to account for the effects of OLS bias is consistent for the true lag order.

3.4. Simplifications of the QFR Estimator

The QFR estimator $\hat{L}_{n,T}(k)$ can be simplified under additional restrictions on the panel process $y_{i,t}$ or approximations based on the relative size of n, T and/or $k_{n,T}$. We consider some leading cases below.

Remarks. (i) *Cross section independence.* If $\mathrm{E}\left(e_{i,t}e'_{j,t}\right) = \mathbf{0}_{m\times m}$ for $i \neq j$, the term $\Pi(k)$ simplifies to $\Pi(k) = \frac{1}{nk}\sum_{i=1}^n \Sigma_{i,i}\,\mathrm{tr}\left(\mathbf{o}_{i,i}(k)\,\Gamma^{-1}(k)\right)$.

[14] We use $o\left(\left\|L_{n,T}(k)\right\|\right)$ to indicate that an arbitrary scalar sequence $b_{n,T}$ satisfies $b_{n,T}/\left\|L_{n,T}(k)\right\| \to 0$.

Then $\hat{\Pi}(k)$ used in the QFR estimator (14) can be simplified as follows:

$$\hat{\Pi}(k) = \frac{1}{n(T-k)k} \sum_{i=1}^{n} \sum_{t=k}^{T-1} \left[\hat{u}_{i,t+1}\left(k, k_{n,T}\right) \hat{u}'_{i,t+1}\left(k, k_{n,T}\right) \operatorname{tr}\left(\ddot{Y}_{i,t}\left(k, k_{n,T}\right) \ddot{Y}'_{i,t}\left(k, k_{n,T}\right) \hat{\mathbf{Q}}^{-1}_{\left(k, k_{n,T}\right)} \right) \right].$$

(ii) *Cross section independence and homoskedasticity.* If $\mathrm{E}\left(e_{i,t} e'_{j,t}\right) = \mathbf{0}_{m \times m}$ for $i \neq j$ and $\Sigma_{i,i} = \Sigma$ for all i, $\Pi(k)$ simplifies to $\Pi(k) = m\Sigma$. In this case the asymptotic expression for the QFR estimator can be simplified as follows:

$$\hat{L}_{n,T}(k) = \hat{R}(k) + \hat{R}(k) \frac{T - k_{n,T}}{T - k_{n,T} - p} \left(\frac{p}{T - k_{n,T}} + \sum_{l=1}^{p} \frac{2l-1}{T-k} \frac{\prod_{r=0}^{l-1}(T-k+r)}{\prod_{r=0}^{l-1}(T-k-r)} + \frac{k}{n\left(T - k_{n,T}\right)} + \frac{k}{n(T-k)} \right) \quad \textbf{(21)}$$
$$+ \hat{\zeta}(k) \left[\frac{kp^2}{\left(T - k_{n,T}\right)^2} + \frac{kp^2}{(T-k)^2} \right].$$

(iii) *Large T relative to $k_{n,T}$.* Note that $(T-k)^{-1} = T^{-1} + kT^{-1}(T-k)^{-1} = T^{-1} + o\left(T^{-1}\right)$, since $k_{n,T} T^{-1} = o(1)$ under Assumption 3. This also implies that

$$\sum_{l=1}^{p} \frac{2l-1}{T-k} \left(\frac{\prod_{r=0}^{l-1}(T-k+r)}{\prod_{r=0}^{l-1}(T-k-r)} - 1 \right) = \frac{p^4 - p^2}{T^2} + o\left((T-k)^{-2} \right).$$

Therefore if T is of moderately large size compared to the maximum lag order $k_{n,T}$, such that $k_{n,T} T^{-1} \simeq 0$, the following simplification of $\hat{L}_{n,T}(k)$ is easier to construct:

$$\hat{L}_{n,T}(k) = \hat{R}(k) + \hat{R}(k) \frac{T - k_{n,T}}{T - k_{n,T} - p} \left(\frac{p}{T - k_{n,T}} + \frac{p^2}{T-k} + \frac{p^4 - p^2}{T^2} \right) + \hat{\Pi}(k) \frac{2k}{nT} + \hat{\zeta}(k) \frac{2kp^2}{T^2}.$$

In the suggested simplifications to follow below in Remarks (iv) and (v), we will employ this $k_{n,T} T^{-1} \simeq 0$ approximation. Note that we could also employ the $k_{n,T} T^{-1} \simeq 0$ approximation in conjunction with the cross section independence and homoskedasticity simplifications given directly above.

(iv) *Large n relative to T.* Suppose that n is substantially larger than T, such that $n^{-1} T \simeq 0$. In this case we could omit the $\hat{\Pi}(k)$ term from the QFR estimator defined in (14), so that

$$\hat{L}_{n,T}(k) = \hat{R}(k) + \hat{R}(k) \frac{T - k_{n,T}}{T - k_{n,T} - p} \left(\frac{p}{T - k_{n,T}} + \frac{p^2}{T-k} + \frac{p^4 - p^2}{T^2} \right) + \hat{\zeta}(k) \frac{2kp^2}{T^2}.$$

Model selection based on the simplified QFR estimator would continue to satisfy Theorem 3.2 provided that $T = o(n)$ as $n \to \infty$ and $T \to \infty$.

(v) *Large T relative to n.* Suppose that T is substantially larger than n, such that $T^{-1} n \simeq 0$. The QFR estimator defined in (14) can then be simplified to $\hat{L}_{n,T}(k) = \hat{R}(k) + \hat{\Pi}(k) \frac{2k}{nT}$. Model selection based on the simplified QFR estimator would continue to satisfy Theorem 3.2 provided that $n = o(T)$ as $n \to \infty$ and $T \to \infty$.

3.5. Bias-Corrected Least Squares

The asymptotic QFR expression (9) shows how the Nickell bias increases the QFL of the OLS fitting. Correcting the LS estimator for this bias before making the forecast therefore presents a potential way to improve forecast accuracy. Bias corrections such as that proposed by Hahn and Kuersteiner (2002) use analytic expressions for the bias in conjunction with the fitted LS parameters to make a first order correction for the $O\left(T^{-1}\right)$ bias of the LS estimator. The correction does not inflate the asymptotic variance of the estimator (Hahn and Kuersteiner, 2002), making BCLS quite attractive from the forecasting perspective. However, the analytic expression that forms the basis of the bias-correction is based on the assumption that the model is correctly specified. In misspecified models, such as those studied here, the correction can exacerbate the bias (Lee, 2006), suggesting that bias correction could in fact inflate QFR in some circumstances.

In this subsection we study BCLS forecasts within the framework described above. First we define the BCLS fitting. Recall that k denotes the fitted lag order and l dictates the beginning of the sample used in estimation (see Remark 2.2 above). Then for all k and l such that $1 \leq k \leq l \leq k_{n,T}$, we define

$$\tilde{\boldsymbol{\alpha}}(k,l) := \hat{\boldsymbol{\alpha}}(k,l) + \tfrac{p}{T-l}\hat{\mathbf{Q}}_{(k,l)}^{-1}\hat{\boldsymbol{\xi}}_{(k,l)}, \tag{22}$$

where $\hat{\boldsymbol{\alpha}}(k,l)$, $\hat{\mathbf{Q}}_{(k,l)}^{-1}$ and $\hat{\boldsymbol{\xi}}_{(k,l)}$ are defined as above in (5), (16) and (19). Note that the order of the quadratic trend p affects the bias correction in the expression above. Next, we define $\tilde{\beta}_i(k,l) := \left(\varsigma_{T-l}'\varsigma_{T-l}\right)^{-1}\varsigma_{T-l}'\left(\mathbf{y}_i^{(l)} - \mathbf{Y}_i^{(k,l)}\tilde{\boldsymbol{\alpha}}(k,l)\right)$, where $\mathbf{y}_i^{(l)} := \left[y_{i,l+1} : \cdots : y_{i,T}\right]'$.

For each $i = 1,\ldots,n$, the BCLS out-of-sample forecast from the fitted VAR(k) is

$$\tilde{y}_{i,T+1}(k) := \tilde{\boldsymbol{\alpha}}(k,k)' Y_{i,T}(k) + \tilde{\beta}_i(k,k)'\tau_{T-k+1},\ 1 \leq k \leq k_{n,T},$$

with an associated quadratic forecast loss of $\mathrm{tr}(\Phi\mathcal{B}(k))$, where

$$\mathcal{B}(k) := \tfrac{1}{n}\sum\nolimits_{i=1}^{n}\left(\tilde{y}_{i,T+1}(k) - y_{i,T+1}\right)\left(\tilde{y}_{i,T+1}(k) - y_{i,T+1}\right)'.$$

The following result characterizes the (second order) asymptotic QFR of the BCLS fitting.

THEOREM 3.3. *Let $y_{i,t}$ be generated according to (1). Then under Assumptions 1, 2, and 3,*

$$\lim_{n,T\to\infty}\max_{1\leq k\leq k_{n,T}}\left|\frac{\mathrm{tr}\left(\Phi\left[\mathrm{E}(\mathcal{B}(k)) - \left(1+\frac{p^2}{T}\right)\Sigma\right]\right)}{\mathrm{tr}\left(\Phi B_{n,T}(k)\right)} - 1\right| = 0$$

for all positive semi definite $m \times m$ matrices Φ satisfying $\|\Phi\| = 1$, and

$$B_{n,T}(k) := \Lambda(k) + \Sigma\left[\frac{kp^2}{T(T-k)} + \sum_{l=1}^{p}\frac{2l-1}{T-k}\left(\frac{\prod_{r=0}^{l-1}(T-k+r)}{\prod_{r=0}^{l-1}(T-k-r)} - 1\right)\right] + \Pi(k)\frac{k}{n(T-k)}.$$

Thus $\mathrm{tr}\left(\Phi B_{n,T}(k)\right)$ is the (second order) asymptotic QFR of the BCLS forecast. Note that $L_{n,T}(k) - B_{n,T}(k) > 0$ (in the matrix sense), indicating that the bias-correction reduces the asymptotic QFR of the LS estimator. Thus, although BC may exacerbate the bias of the estimator under model misspecification (Lee, 2006), the Theorem reveals an asymptotic justification for the use of bias-correction. Intuitively, in severely misspecified models (i.e., parsimonious models), the bias is dominated by model misspecification in the asymptotic QFR expression, so that the $O\left(T^{-1}\right)$ bias is of second-order concern. In this situation, any exacerbation of the bias from implementing the bias correction is dominated by the effect of misspecification on QFR. In mildly misspecified models (i.e., larger models), the effect of misspecification on the asymptotic QFR is much smaller. However, the correction becomes more accurate under mild misspecification, and so it improves forecast accuracy in sufficiently large samples.

In order to select the lag order of the bias-corrected forecasting model we minimize a modified panel FPE estimator that follows the same structure as (14). First we define $\tilde{u}_{i,t+1}(k,l) := y_{i,t} - \tilde{\alpha}(k,l)' Y_{i,t}(k) - \tilde{\beta}_i(k,l)' \tau_{t-l+1}$ for $i = 1,\ldots,n$ and $t = l,\ldots,T$ as the vector of residuals from the BCLS fitting. We then use

$$\hat{B}_{n,T}(k) := \tilde{R}(k) + \tilde{R}(k)\frac{T-k_{n,T}}{T-k_{n,T}-p}\left(\frac{p}{T-k_{n,T}} + \sum_{l=1}^{p}\frac{2l-1}{T-k}\frac{\prod_{r=0}^{l-1}(T-k+r)}{\prod_{r=0}^{l-1}(T-k-r)}\right) + \tilde{\Pi}(k)\frac{k}{n}\left[\frac{1}{T-k_{n,T}} + \frac{1}{T-k}\right]$$

as the estimator of the QFR, where $\tilde{R}(k)$ and $\tilde{\Pi}(k)$ are defined as above in (13) and (17), but with $\tilde{u}_{i,t+1}\left(k,k_{n,T}\right)$ replacing $\hat{u}_{i,t+1}\left(k,k_{n,T}\right)$ in the formulae. Note that many of the simplifications of the QFR estimator described in Section 3.4 can be applied to $\hat{B}_{n,T}(k)$.

In the Appendix we show that $\hat{B}_{n,T}(k)$ is an unbiased estimator of the QFR of the BCLS fitting (Theorem B.10), which provides a theoretical motivation for a lag selection criterion of the form

$$\tilde{k} = \arg\min_{1\le k\le k_{n,T}} \mathrm{tr}\left(\Phi \hat{B}_{n,T}(k)\right). \tag{23}$$

Theorem 3.4 below establishes the asymptotic efficiency of the model selection rule.

THEOREM 3.4. *Let $y_{i,t}$ be generated according to (1), and assume that there exists $\upsilon \in (0,1)$ such that either $n^{\upsilon-1}T = o(1)$ or $T^{\upsilon-1}n = o(1)$ as $n \to \infty$ and $T \to \infty$. Then under Assumptions 1, 2, 3, and 4,*

$$\lim_{n,T\to\infty} \frac{\mathrm{tr}\left(\Phi\left[\mathrm{E}\left(\mathcal{B}\left(\tilde{k}\right)\right) - \left(1+\frac{p^2}{T}\right)\Sigma\right]\right)}{\mathrm{tr}\left(\Phi B_{n,T}\left(k_{n,T}^{**}\right)\right)} = 1,$$

for all positive semidefinite $m \times m$ matrices Φ satisfying $\|\Phi\| = 1$, and where

$$k_{n,T}^{**} := \arg\min_{1\le k\le k_{n,T}} \mathrm{tr}\left(\Phi B_{n,T}(k)\right).$$

Because $\mathrm{tr}\left(\Phi B_{n,T}\left(k_{n,T}^{**}\right)\right) \leq \mathrm{tr}\left(\Phi B_{n,T}\left(k_{n,T}^{*}\right)\right) \leq \mathrm{tr}\left(\Phi L_{n,T}\left(k_{n,T}^{*}\right)\right)$, Theorem 3.4 implies it is possible to significantly enhance forecast accuracy in large samples by using a BCLS forecast selected using (23) instead of employing a LS forecast. In addition, because (i) $\mathrm{tr}\left(\left(L_{n,T}(k) - B_{n,T}(k)\right)\Phi\right) = \frac{kp^2}{(T-k)^2}\mathrm{tr}\left(\Phi\zeta(k)\right) > 0$ is increasing in k (see Remark (iv) in Section 3.1), and (ii) $\mathrm{tr}\left(B_{n,T}(k)\Phi\right) > 0$ and $\mathrm{tr}\left(L_{n,T}(k)\Phi\right) > 0$ are both convex in k, it follows that $k_{n,T}^{*} \leq k_{n,T}^{**}$. Thus, in large samples, the selected BCLS forecast exhibits less specification error compared to the selected LS forecast.

4. MONTE CARLO STUDIES

We conduct Monte Carlo studies in order to verify the theoretical results given above and to investigate the finite sample properties of the proposed panel FPE model selection criterion. We fit a set of AR(k) models to an ARMA(1, 1) process of the general form

$$y_{i,t} = 0.8 \cdot y_{i,t-1} + \rho e_{i,t-1} + e_{i,t},\ i = 1,\ldots,n;\ t = 2,\ldots,T+1. \tag{24}$$

Misspecification of the set of AR(k) models occurs when $\rho \neq 0$. To generate weak cross sectional dependence, $e_{i,t}$ follows

$$e_{i,t} = \phi e_{i-1,t} + u_{i,t},\ u_{i,t} \sim iidN\left(0, \left(1-\phi^2\right)^{-1/2}\right). \tag{25}$$

Note that $e_{i,t}$ has unit variance. For each simulated panel we fit AR(k) models for each $k = 1,\ldots,k_{n,T}$. We construct both the quadratic forecast loss (QFL) of the forecasts $\left\{\hat{y}_{i,T+1}(k)\right\}_{i=1}^{n}$ and the estimated QFR for each fitted AR(k) model. The maximum lag order $k_{n,T}$ is determined by the sample dimensions of the simulated panel. The fitted models include cross section fixed effects ($p = 1$) or heterogenous linear trends ($p = 2$). Although these cross sectional effects are omitted from the data generating process (DGP), including them in the fitted models will have the desired effect on the simulated QFLs. In our finite sample investigation we also compute quantities which form the basis for other lag selection criteria, as we will outline below in more detail. Each simulation is replicated 4,000 times. Tables containing the simulation results are located in Appendix A.

4.1. Verifying Asymptotic Expressions

In this section we discuss the results of some large sample simulations in order to verify the Theorems presented above. In all simulations we set $\phi = 0.5$ so that the time series exhibit weak cross section dependence.

Large n Simulations. We consider $T = 25, 50, 100, 200$, $n = \mathrm{int}(0.1 \times T^2)$, and set $k_{n,T} = 0.2 \times T$, so that $k_{n,T} = O(T)$ (which is just above the maximum permissible growth rate). We can consider larger $k_{n,T}$, but the computational time quickly becomes burdensome due to the large lag lengths permitted and the large size of the panels considered.

In Table A1 we compare the average of the simulated (second order) QFLs to the analytic expression for $L_{n,T}(k)$. Specifically, we tabulate $\max_{1\le k\le k_{n,T}}\left|L_{n,T}^{-1}(k)\bar{\mathcal{L}}(k)-1\right|$, where $\bar{\mathcal{L}}(k)$ denotes the empirical average of $\mathcal{L}(k)-\left(1+p^2T^{-1}\right)$ across the 4,000 replications. The table entries evidently approach zero as T grows large, in accordance with Theorem 3.1. Table A3 exhibits a similar result for the BCLS forecasts, in accordance with Theorem 3.3.

In order to demonstrate asymptotic efficiency we follow Ing and Wei (2005) and compare the average simulated QFLs selected by panel FPE-minimization to $L_{n,T}\left(k_{n,T}^*\right)$. Specifically, in Table A2 we exhibit $\max_{1\le k\le k_{n,T}}\left|L_{n,T}^{-1}\left(k_{n,T}^*\right)\bar{\mathcal{L}}\left(\hat{k}\right)\right|$, where $\bar{\mathcal{L}}\left(\hat{k}\right)$ denotes the empirical average of $\mathcal{L}(\hat{k})-\left(1+p^2T^{-1}\right)$ across the 4,000 replications. Under Theorem 3.2 the ratios should approach one as the sample size grows large. In general, the ratios reported in the table are larger than one. But as T and n grow large, there is generally a convergence towards unity, consistent with the Theorem. Table A4 exhibits similar results for the lag order selection method tailored to BCLS forecasting given in (23).

Large T Simulations. We consider $n=5,10,20,40$, $T=n^2$, and set $k_{n,T}=n$, so that $k_{n,T}=O\left(\sqrt{T}\right)$, which is just above the maximum permissible rate of growth.

Tables A5 and A6 exhibit the same quantities as those given in Tables A1 and A2, respectively. The table entries in Table A5 approach zero from above as the sample size grow large, in accordance with Theorem 3.1. Meanwhile, the entries in Table A6 approach one from above as the sample size grows large, in accordance with Theorem 3.2, although the decline is not monotonic. The table entries are much larger than those exhibited in Table 3.2 (perhaps due to the fewer number of observations in these simulated samples) but they are comparable to the results exhibited in Table 1 of Ing and Wei (2005) for the time series case. Tables A7 and A8 exhibit similar results for the bias-corrected LS forecasts and associated lag order selection method defined in (23).

4.2. Finite Sample Performance of Model Selection Criteria

In this subsection we compare the proposed model selection criteria to some alternative methods. It is important to note that the alternatives are not tailored to the problem of forecasting a VAR(∞), and thus we expect our new criteria to perform well compared to these alternatives.

We consider eight different DGPs by considering all combinations of $\rho\in\{0.7,0.3,-0.3,-0.7\}$ and $\phi\in\{0.5,0\}$. In terms of sample size, we consider all combinations of $n\in\{10,15,25,50,100,200\}$ and $T\in\{10,15,25,50\}$.

The maximum permissible lag order $k_{n,T}$ is set to the integer closest to $\min\left(2\times\max\left(\sqrt{n},\sqrt{T}\right),\frac{1}{2}T\right)$. For brevity we will consider only the fixed effects case ($p=1$) here.

Conventional Final Prediction Error. One of our main contributions is to show the importance of penalizing bias when selecting a LS forecast. We therefore compare our suggested criterion (15) to model selection by FPE-minimization originally proposed by Akaike (1970), which penalizes only the variance of the fitting. This "conventional FPE" criterion is

$$\hat{k}^{\mathrm{FP}} := \arg\min_{1\le k\le k_{n,T}}\left[\hat{R}(k)+\frac{T-k_{n,T}}{T-k_{n,T}-p}\hat{R}(k)\left(\frac{1}{T-k}+\frac{1}{T-k_{n,T}}\right)+\hat{\Pi}(k)\left(\frac{k}{n(T-k)}+\frac{k}{n(T-k_{n,T})}\right)\right]. \tag{26}$$

This criterion (26) might be constructed by a practitioner who knows that the variance of the fitted model should be penalized, but ignores the effect of bias.

Table A9 reports the ratio of the average QFL selected using (15) to the average QFL selected by the conventional FPE criterion defined in (26). The table entries are generally less than one, indicating that the panel FPE criterion outperforms the conventional FPE on average in most simulations considered. However, when $\rho=0.7$, the conventional FPE tends to perform better when $n\ge 25$ and $T\le 25$. This may be because the estimates of quadratic bias required for the panel FPE are less accurate under severe model misspecification.

Kullback–Leibler Information Criterion. The importance of penalizing $O\left(T^{-1}\right)$ bias in panel model selection has been noted recently by Lee and Phillips (2015). The Lee and Phillips Kullback–Leibler information criterion (KLIC) is a potential alternative to the panel FPE proposed here since model selection via minimizing Gaussian Kullback–Leibler information loss (i.e., model selection by AIC) generates asymptotically efficient forecasts in the time series context (Shibata, 1980; Ing and Wei, 2005). We use the exact expression for the KLIC (see p. 9 of Lee and Phillips, 2015):

$$\hat{k}^{\mathrm{KLIC}} := \arg\min_{1\le k\le k_{n,T}}\left[\ln\left(\tilde{R}(k)\left\{1+\frac{1}{T-k_{n,T}}\tilde{R}^{-1}(k)\tilde{V}^{\circ}(k)\right\}\right)+\frac{2k}{n(T-k_{n,T})}\right], \tag{27}$$

where $\tilde{V}^{\circ}(k)$ denotes a kernel-based estimator of the long-run (time series) variance applied to the regression residuals from the BCLS fitting averaged over cross sectional units. As in the Monte Carlo study of Lee and Phillips (2015), we only use the residuals that span $t=k_{n,T}+1,\ldots,T$ to construct $\tilde{V}^{\circ}(k)$ (and $\tilde{R}(k)$), and the residuals are based on bias-corrected least-squares estimates (specifically we use the Hahn and Kuersteiner (2002) bias correction defined in (22)). Including the long-run correlation term $\tilde{R}^{-1}(k)\tilde{V}^{\circ}(k)$ in the log function effectively penalizes the effect of LS bias, because the within transformation used to remove incidental parameters induces serial correlation in the regression residuals. To construct $\tilde{V}^{\circ}(k)$ we use a Bartlett kernel with the bandwidth set to the largest integer less than $0.75\left(T-k_{n,T}\right)^{\frac{1}{3}}$.[15]

Table A10 reports the ratio of the average QFL selected using the panel FPE criterion defined in (15) to the average QFL selected by KLIC defined in (27). The FPE tends to perform better than KLIC across the majority of simulations. Note, however, that the KLIC performs better when the moving average parameter is rather large ($\rho = 0.7$) and the time series dimension of the panel is limited ($T \leq 15$). As previously discussed, this is perhaps due to the fact that the misspecification of the autoregressive model is rather severe for this DGP, and it adversely affects the accuracy of our QFR estimator in moderate sample sizes.

The panel FPE outperforms the KLIC in the forecasting exercise because the latter is designed to minimize the information loss associated with the homogenous parameters of the model *after* integrating out the incidental parameters. For the purposes of forecasting, the information loss of the fitted heterogenous parameters (such as cross section fixed effects) affects forecast accuracy. The panel FPE imposes an additional overfitting penalty in order to account for the effect of variability in these estimated fixed effects. Also, the KLIC is designed under the assumption of cross section independence, and hence it is likely to underfit the model in the presence of positive correlation between cross sectional units.

Bias-Corrected Least Squares. We also explore the finite sample performance of BCLS forecast relative to the LS forecast. The BCLS forecasting model is selected using the criterion defined in (23), while the LS forecasting model is selected using the method defined in (15). Table A11 exhibits the ratio of the average QFL of the selected LS forecasts to the average QFL of the selected BCLS forecasts. The table entries are almost exclusively greater than one, indicating that the BCLS forecast generally outperforms the LS forecast. This suggests that bias-correction is a viable method to significantly improve LS accuracy in practice, even in moderate or small sample sizes. Table A12 compares the average lag order selected for the LS and BCLS forecasts, and shows that the bias-corrected panel FPE criterion selects a larger lag order on average in almost all simulation designs, as expected.

5. SUMMARY

In this article we suggest new model selection methods for panel data VAR forecasting. Following Akaike (1969, 1970), the model selection method is based on minimizing the estimated out-of-sample quadratic forecast risk. Relative to comparable time series forecast selection criteria, the panel criterion places a larger

[15] We use the bandwidth rule suggested by Stock and Watson (2010, p. 641) for processes with moderate amounts of serial dependence. We also considered a larger truncation set to grow at a rate slightly slower than $\left(T-k_{n,T}\right)^{1/2}$ (this is largest permissible truncation lag permitted; see the discussion of p. 7 of Lee and Phillips, 2015). The performance of the KLIC improved slightly in some simulations and was worse in others.

penalty on model complexity in order to account for the effects of parameterized heterogeneity on forecast loss. We demonstrate that the selection criterion is asymptotically efficient in the sense of Shibata (1980) and Ing and Wei (2005) under fairly general conditions. We extend the analysis to bias-corrected least squares forecasts, showing that an analogous model selection criterion based on QFR-minimization is also asymptotically efficient.

Several extensions to this research appear promising. First, we have focussed only on one-step-ahead prediction. A generalization to multistep forecasting in the vein of Bhansali (1996, 1997) would benefit practitioners. As shown in Greenaway-McGrevy (2013), the bias function becomes more complicated in correctly specified multistep panel regressions, and these complications will be further exacerbated once we permit model misspecification. Second, we established asymptotic efficiency under conditions that preclude n and T growing at the same rate. Methods applicable to large n and T panels should be explored. Third, we have not considered how to select the order of the trend polynomial p. The results derived herein can however inform this selection problem: Theorem 3.1 shows that the polynomial trend introduces an $O\left(p^2T^{-1}\right)$ term into the QFR, indicating that overfitting the trend order can entail a substantial reduction in forecast accuracy. Fourth, while our focus has been on selecting a single model for prediction, forecast combination can yield substantial improvements in the forecast accuracy of time series models. Using the asymptotic expressions provided herein to generalize this approach to the panel forecasting framework could potentially yield further improvements. These topics are left for future research.

REFERENCES

Akaike, H. (1970) Statistical predictor identification. *Annals of the Institute of Statistical Mathematics* 22, 203–419.

Alvarez, J. & M. Arellano (2003) The time series and cross-section asymptotics of dynamic panel data estimators. *Econometrica* 71, 1121–1159.

Baillie, R.T. & B.H. Baltagi (1999) Prediction from the regression model with one-way error components. In C. Hsiao, K. Lahiri, L.F. Lee, & H. Pesaran (eds.), *Analysis of Panels and Limited Dependent Variable Models*, pp. 255–267. Cambridge University Press.

Baltagi, B.H. (2008) Forecasting with panel data. *Journal of Forecasting* 27, 153–173.

Baltagi, B.H., G. Bresson, & A. Pirotte (2002) Comparison of forecast performance for homogeneous, heterogeneous and shrinkage estimators: Some empirical evidence from US electricity and natural-gas consumption. *Economics Letters* 76, 375–382.

Baltagi, B.H. & J.M. Griffin (1997) Pooled estimators vs. their heterogeneous counterparts in the context of dynamic demand for gasoline. *Journal of Econometrics* 77, 303–327.

Baltagi, B.H., J.M. Griffin, & W. Xiong (2000) To pool or not to pool: Homogeneous versus heterogeneous estimators applied to cigarette demand. *Review of Economics and Statistics* 82, 117–126.

Baltagi, B.H. & Q. Li (1992) Prediction in the one-way error component model with serial correlation. *Journal of Forecasting* 11, 561–567.

Bhansali, R.J. (1996) Asymptotically efficient autoregressive model selection for multistep prediction. *Annals of the Institute of Statistical Mathematics* 48, 577–602.

Bhansali, R.J. (1997) Direct autoregressive predictors for multistep prediction: Order selection and performance relative to plug-in predictors. *Statistica Sinica* 7, 425–449.

Brillinger, D.R. (1981) *Time Series*: *Data Analysis and Theory*. Holden-Day.

Brucker H. & B. Siliverstovs (2006) On the estimation and forecasting of international migration: How relevant is heterogeneity across countries? *Empirical Economics* 31, 735–754.

Chudik, A., M.H. Pesaran, & E. Tosetti (2011) Weak and strong cross-section dependence and estimation of large panels. *The Econometrics Journal*, 1, 45–90.

Driver, C., K. Imai, P. Temple, & A. Urga (2004) The effect of uncertainty on UK investment authorisation: Homogeneous vs. heterogeneous estimators. *Empirical Economics* 29, 115–128.

Eisinberg, A. & G. Fedele (2007) Discrete orthogonal polynomials on equidistant nodes. *International Mathematical Forum* 21, 1007–1020

Findley, D.F. (1984) On some ambiguities associated with the fitting of ARMA models to time series. *Journal of Time Series Analysis* 5, 213–225.

Findley, D.F. & C.Z. Wei (1993) Moment bounds for deriving time series CLT's and model selection procedures. *Statistica Sinica* 3, 453–480.

Findley, D.F. & C.Z. Wei (2002) AIC, overfitting principles, and boundedness of moments of inverse matrices for vector autoregressions and related models. *Journal of Multivariate Analysis* 83, 415–450.

Fuller, W.A. & D.P. Hasza (1981) Properties of predictors for autoregressive time series. *Journal of American Statistical Association* 76, 155–161.

Gavin W.T. & A.T. Theodorou (2005) A common model approach to macroeconomics: Using panel data to reduce sampling error. *Journal of Forecasting* 24, 203–219.

Greenaway-McGrevy, R. (2013) Multistep prediction of panel autoregressive processes. *Econometric Theory* 29, 699–734

Greenaway-McGrevy, R. (2015) Evaluating panel data forecasts under independent realization. *Journal of Multivariate Analysis* 136, 108–125.

Hahn, J. & G. Kuersteiner (2002) Asymptotically unbiased inference for a dynamic panel model with fixed effects when both n and T are large. *Econometrica* 70, 1639–1657.

Hahn, J. & G. Kuersteiner (2011) Bias reduction for dynamic nonlinear panel models with fixed effects. *Econometric Theory* 27, 1152–1191.

Han, C., P.C.B. Phillips, & D. Sul (2017) Lag length selection in panel autoregression. *Econometric Reviews* 36, 225–240.

Hansen, C.B. (2007) Asymptotic properties of a robust variance matrix estimator for panel data when T is large. *Journal of Econometrics* 141, 597–620.

Hoogstrate, A.J., F.C. Palm, & G.A. Pfann (2000) Pooling in dynamic panel-data models: An application to forecasting GDP growth rates. *Journal of Business and Economic Statistics* 18, 274–283.

Ing, C.K. (2003) Multistep prediction in autoregressive processes. *Econometric Theory* 19, 254–279.

Ing, C.K. & C.Z. Wei (2003) On same-realization prediction in an infinite-order autoregressive process. *Journal of Multivariate Analysis* 85, 130–155.

Ing, C.K. & C.Z. Wei (2005) Order selection for same-realization predictions in autoregressive processes. *Annals of Statistics* 33, 2423–2474.

Kiviet, J.F. (1995) On bias, inconsistency and efficiency of some estimators in dynamic panel data models. *Journal of Econometrics* 68, 53–78.

Kunitomo, N. & T. Yamamoto (1985) Properties of predictors in misspecified autoregressive time series models. *Journal of the American Statistical Association* 80, 941–950.

Lee, Y. (2006) A General Approach to Bias Correction in Dynamic Panels under Time Series Misspecification. Ph.D. dissertation, Yale University.

Lee, Y. (2012) Bias in dynamic panel models under time series misspecification. *Journal of Econometrics* 169, 54–60.

Lee, Y. & P.C.B. Phillips (2015) Model selection in the presence of incidental parameters. *Journal of Econometrics* 188, 474–489.

Mallows, C.L. (1973) Some comments on C_p. *Technometrics* 15, 661–675.

Mark, N. & D. Sul (2001) Nominal exchange rates and monetary fundamentals: Evidence from a seventeen country panel. *Journal of International Economics* 53, 29–52.

Nickell, S. (1981) Biases in dynamic models with fixed effects. *Econometrica* 49, 1417–1425.

Phillips, P.C.B. & D. Sul (2007) Bias in dynamic panel estimation with fixed effects, incidental trends and cross section dependence. *Journal of Econometrics* 137, 162–188.
Rapach, D.E. & M.E. Wohar (2004) Testing the monetary model of exchange rate determination: A closer look at panels. *Journal of International Money and Finance* 23, 841–865.
Schorfheide, F. (2005) VAR forecasting under misspecification. *Journal of Econometrics* 128, 99–136.
Shibata, R. (1980) Asymptotically efficient selection of the order of the model for estimating parameters of a linear process. *Annals of Statistics* 8, 147–164.
Speed, T.P. & B. Yu (1993) Model selection and prediction: Normal regression. *Annals of the Institute of Statistical Mathematics* 45, 35–54.
Stock, J.H. & M.W. Watson (2010) *Introduction to Econometrics*, 3rd ed. Addison-Wesley.
Taub, A.J. (1979) Prediction in the context of the variance components model. *Journal of Econometrics* 10, 103–107.

APPENDIX A: Tables

TABLE A1. Asymptotic QFR (Theorem 3.1)

n	T	$k_{n,T}$	$\rho=0.7$	$\rho=0.5$	$\rho=0.3$	$\rho=0.1$	$\rho=-0.1$	$\rho=-0.3$	$\rho=-0.5$	$\rho=-0.7$
Cross section fixed effects ($p=1$)										
63	25	5	0.10	0.17	0.18	0.17	0.17	0.16	0.19	0.32
250	50	10	0.06	0.06	0.06	0.05	0.05	0.05	0.05	0.10
1,000	100	20	0.02	0.02	0.02	0.02	0.01	0.01	0.01	0.03
4,000	200	40	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Heterogenous linear trends ($p=2$)										
63	25	5	0.24	0.29	0.30	0.29	0.31	0.31	0.40	0.53
250	50	10	0.13	0.12	0.12	0.12	0.12	0.11	0.12	0.26
1,000	100	20	0.05	0.05	0.05	0.04	0.04	0.04	0.04	0.08
4,000	200	40	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02

Table entries are $\max_{1\le k\le k_{n,T}}\left|L_{n,T}^{-1}(k)\,\bar{\mathcal{L}}(k)-1\right|$, where $\bar{\mathcal{L}}(k)$ denotes the empirical average of $\mathcal{L}(k)-\left(1+p^2T^{-1}\right)$ across the 4,000 replications.

TABLE A2. Asymptotic efficiency of panel FPE minimization (Theorem 3.2)

n	T	$k_{n,T}$	$\rho=0.7$	$\rho=0.5$	$\rho=0.3$	$\rho=0.1$	$\rho=-0.1$	$\rho=-0.3$	$\rho=-0.5$	$\rho=-0.7$
Cross section fixed effects ($p=1$)										
63	25	5	0.98	0.89	1.01	1.14	1.03	1.34	1.25	0.71
250	50	10	1.03	1.11	1.31	1.18	1.51	1.21	1.18	1.49
1,000	100	20	1.03	1.02	1.02	1.00	1.00	0.99	1.14	1.39
4,000	200	40	1.01	1.09	1.01	1.04	1.01	1.32	1.15	1.28
Heterogenous linear trends ($p=2$)										
63	25	5	0.78	0.87	0.85	1.12	0.86	0.77	0.67	0.55
250	50	10	1.01	1.00	1.02	1.13	0.97	1.05	1.17	0.81
1,000	100	20	1.03	1.06	1.05	0.97	1.23	1.26	1.12	1.49
4,000	200	40	1.01	1.00	0.99	0.99	0.99	1.01	1.15	1.30

Table entries are $\left|L_{n,T}^{-1}\left(k_{n,T}^{*}\right)\bar{\mathcal{L}}\left(\hat{k}\right)\right|$, where $\bar{\mathcal{L}}\left(\hat{k}\right)$ denotes the empirical average of $\mathcal{L}(\hat{k})-\left(1+p^2T^{-1}\right)$ across the 4,000 replications, and $\hat{k}$ denotes the lag order selected by panel FPE minimization (15).

TABLE A3. Asymptotic QFR for BCLS (Theorem 3.3)

n	T	$k_{n,T}$	$\rho=0.7$	$\rho=0.5$	$\rho=0.3$	$\rho=0.1$	$\rho=-0.1$	$\rho=-0.3$	$\rho=-0.5$	$\rho=-0.7$
					Cross section fixed effects ($p=1$)					
63	25	5	0.29	0.27	0.20	0.41	0.26	0.28	0.32	0.44
250	50	10	0.09	0.08	0.10	0.12	0.08	0.13	0.17	0.21
1,000	100	20	0.04	0.06	0.06	0.04	0.05	0.07	0.07	0.10
4,000	200	40	0.02	0.03	0.04	0.06	0.05	0.03	0.03	0.04
					Heterogenous linear trends ($p=2$)					
63	25	5	0.36	0.42	0.48	0.25	0.57	0.49	0.61	0.71
250	50	10	0.16	0.17	0.20	0.26	0.25	0.30	0.32	0.39
1,000	100	20	0.07	0.08	0.07	0.12	0.10	0.14	0.16	0.20
4,000	200	40	0.04	0.04	0.05	0.04	0.04	0.06	0.07	0.10

Table entries are $\max_{1\le k\le k_{n,T}}\left|B_{n,T}^{-1}(k)\bar{\mathcal{B}}(k)-1\right|$, where $\bar{\mathcal{B}}(k)$ denotes the empirical average of $\mathcal{B}(k)-\left(1+p^2T^{-1}\right)$ across the 4,000 replications.

TABLE A4. Asymptotic efficiency of panel FPE minimization for BCLS (Theorem 3.4)

n	T	$k_{n,T}$	$\rho=0.7$	$\rho=0.5$	$\rho=0.3$	$\rho=0.1$	$\rho=-0.1$	$\rho=-0.3$	$\rho=-0.5$	$\rho=-0.7$
					Cross section fixed effects ($p=1$)					
63	25	5	1.36	1.00	1.20	1.53	1.45	1.68	1.79	0.83
250	50	10	1.00	1.25	1.57	1.98	2.62	1.87	1.39	2.07
1,000	100	20	1.09	1.15	1.29	1.37	1.30	1.22	1.42	1.76
4,000	200	40	1.05	1.12	1.14	1.51	1.36	1.42	1.50	1.65
					Heterogenous linear trends ($p=2$)					
63	25	5	0.84	0.89	0.78	1.16	0.73	0.98	0.82	0.38
250	50	10	0.96	1.03	1.37	1.63	1.37	0.88	1.26	1.02
1,000	100	20	1.09	1.14	1.21	1.18	1.14	1.84	1.35	1.93
4,000	200	40	1.02	0.99	0.98	0.96	0.96	1.18	1.45	1.79

Table entries are $\left|B_{n,T}^{-1}\left(k_{n,T}^{**}\right)\bar{\mathcal{B}}\left(\tilde{k}\right)\right|$, where $\bar{\mathcal{B}}\left(\tilde{k}\right)$ denotes the empirical average of $\mathcal{B}\left(\tilde{k}\right)-\left(1+p^2T^{-1}\right)$ across the 4,000 replications, and $\tilde{k}$ denotes the lag order selected by BCLS panel FPE minimization (23).

TABLE A5. Asymptotic QFR (Theorem 3.1)

n	T	$k_{n,T}$	$\rho = 0.7$	$\rho = 0.5$	$\rho = 0.3$	$\rho = 0.1$	$\rho = -0.1$	$\rho = -0.3$	$\rho = -0.5$	$\rho = -0.7$
Cross section fixed effects ($p = 1$)										
25	25	5	0.12	0.09	0.10	0.08	0.04	0.07	0.15	0.25
50	100	10	0.02	0.03	0.03	0.03	0.03	0.02	0.03	0.04
100	400	20	0.04	0.04	0.05	0.05	0.05	0.05	0.04	0.04
200	1,600	40	0.02	0.02	0.02	0.05	0.04	0.02	0.02	0.02
Heterogenous linear trends ($p = 2$)										
25	25	5	0.19	0.15	0.07	0.10	0.21	0.25	0.36	0.36
50	100	10	0.06	0.06	0.04	0.04	0.04	0.05	0.05	0.10
100	400	20	0.02	0.05	0.02	0.03	0.04	0.06	0.05	0.03
200	1,600	40	0.04	0.06	0.06	0.10	0.10	0.07	0.06	0.05

Table entries are $\max_{1 \le k \le k_{n,T}} \left| L_{n,T}^{-1}(k)\, \bar{\mathcal{L}}(k) - 1 \right|$, where $\bar{\mathcal{L}}(k)$ denotes the empirical average of $\mathcal{L}(k) - \left(1 + p^2 T^{-1}\right)$ across the 4,000 replications.

TABLE A6. Asymptotic efficiency of panel FPE minimization (Theorem 3.2)

n	T	$k_{n,T}$	$\rho = 0.7$	$\rho = 0.5$	$\rho = 0.3$	$\rho = 0.1$	$\rho = -0.1$	$\rho = -0.3$	$\rho = -0.5$	$\rho = -0.7$
Cross section fixed effects ($p = 1$)										
5	25	5	1.23	1.37	1.61	2.05	1.88	1.46	1.30	1.62
10	100	10	1.16	1.31	1.68	2.01	2.05	1.54	1.40	1.36
20	400	20	1.19	1.33	1.53	1.91	1.93	1.47	1.37	1.31
40	1,600	40	1.18	1.31	1.47	1.91	1.85	1.47	1.33	1.25
Heterogenous linear trends ($p = 2$)										
5	25	5	1.17	1.23	1.37	1.72	1.58	1.17	1.01	1.65
10	100	10	1.20	1.31	1.57	1.71	1.72	1.41	1.42	1.22
20	400	20	1.21	1.33	1.48	1.73	1.76	1.44	1.36	1.31
40	1,600	40	1.21	1.32	1.45	1.86	1.81	1.46	1.34	1.26

Table entries are $\left| L_{n,T}^{-1}\left(k_{n,T}^{*}\right) \bar{\mathcal{L}}\left(\hat{k}\right) \right|$, where $\bar{\mathcal{L}}\left(\hat{k}\right)$ denotes the empirical average of $\mathcal{L}(\hat{k}) - \left(1 + p^2 T^{-1}\right)$ across the 4,000 replications, and $\hat{k}$ denotes the lag order selected by panel FPE minimization (15).

TABLE A7. Asymptotic QFR for BCLS (Theorem 3.3)

n	T	$k_{n,T}$	$\rho = 0.7$	$\rho = 0.5$	$\rho = 0.3$	$\rho = 0.1$	$\rho = -0.1$	$\rho = -0.3$	$\rho = -0.5$	$\rho = -0.7$
					Cross section fixed effects ($p = 1$)					
5	25	5	2.78	4.68	0.60	0.68	0.57	0.47	0.32	0.25
10	100	10	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.05
20	400	20	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.03
40	1,600	40	0.02	0.02	0.03	0.06	0.05	0.03	0.02	0.02
					Heterogenous linear trends ($p = 2$)					
5	25	5	1.39	1.24	1.14	1.07	1.00	0.90	0.69	0.46
10	100	10	0.17	0.16	0.16	0.15	0.15	0.15	0.14	0.11
20	400	20	0.04	0.05	0.04	0.07	0.07	0.08	0.04	0.04
40	1,600	40	0.04	0.06	0.07	0.13	0.13	0.08	0.07	0.05

Table entries are $\max_{1 \le k \le k_{n,T}} \left| B_{n,T}^{-1}(k)\,\bar{\mathcal{B}}(k) - 1 \right|$, where $\bar{\mathcal{B}}(k)$ denotes the empirical average of $\mathcal{B}(k) - \left(1 + p^2 T^{-1}\right)$ across the 4,000 replications.

TABLE A8. Asymptotic efficiency of panel FPE minimization for BCLS (Theorem 3.4)

n	T	$k_{n,T}$	$\rho = 0.7$	$\rho = 0.5$	$\rho = 0.3$	$\rho = 0.1$	$\rho = -0.1$	$\rho = -0.3$	$\rho = -0.5$	$\rho = -0.7$
					Cross section fixed effects ($p = 1$)					
5	25	5	1.52	1.75	2.08	2.80	2.63	1.90	1.57	1.66
10	100	10	1.19	1.35	1.83	2.24	2.39	1.70	1.56	1.49
20	400	20	1.20	1.34	1.56	2.05	2.07	1.55	1.44	1.37
40	1,600	40	1.18	1.32	1.50	1.99	1.93	1.51	1.37	1.27
					Heterogenous linear trends ($p = 2$)					
5	25	5	1.69	1.74	2.12	2.89	2.42	1.56	1.26	1.57
10	50	10	1.28	1.40	1.90	2.15	2.33	1.72	1.65	1.48
20	100	20	1.25	1.39	1.58	2.04	2.08	1.62	1.49	1.43
40	200	40	1.22	1.36	1.55	2.06	2.03	1.56	1.42	1.32

Table entries are $\left| B_{n,T}^{-1}\left(k_{n,T}^{**}\right) \bar{\mathcal{B}}\left(\tilde{k}\right) \right|$, where $\bar{\mathcal{B}}\left(\tilde{k}\right)$ denotes the empirical average of $\mathcal{B}\,(\tilde{k}) - \left(1 + p^2 T^{-1}\right)$ across the 4,000 replications, and $\tilde{k}$ denotes the lag order selected by BCLS panel FPE minimization (23).

TABLE A9. Comparison of panel FPE forecast selection to conventional FPE forecast selection

T	$n=10$	$n=15$	$n=25$	$n=50$	$n=100$	$n=200$	$n=10$	$n=15$	$n=25$	$n=50$	$n=100$	$n=200$
			$\rho=-0.7,\phi=0$						$\rho=-0.3,\phi=0$			
10	0.89	0.88	0.86	0.81	0.75	0.70	0.89	0.92	0.92	0.93	0.96	0.98
15	0.85	0.83	0.80	0.74	0.69	0.64	0.86	0.88	0.90	0.92	0.94	0.97
25	0.86	0.84	0.79	0.72	0.66	0.63	0.97	0.98	1.00	1.00	1.03	1.06
50	0.94	0.94	0.97	1.00	0.88	0.72	0.98	0.95	0.92	0.88	0.74	0.67
			$\rho=0.7,\phi=0$						$\rho=0.3,\phi=0$			
10	0.99	1.03	1.11	1.30	1.30	1.38	0.89	0.93	0.96	0.98	0.99	1.00
15	0.98	1.02	1.05	1.04	1.04	1.04	0.88	0.90	0.93	0.94	0.94	0.96
25	0.99	0.99	1.00	1.01	1.01	1.00	0.95	0.95	0.93	0.90	0.87	0.85
50	0.99	0.98	0.98	0.96	0.90	0.86	0.98	0.96	0.94	0.92	0.81	0.77
			$\rho=-0.7,\phi=0.5$						$\rho=-0.3,\phi=0.5$			
10	0.88	0.89	0.87	0.85	0.79	0.74	0.86	0.89	0.92	0.93	0.94	0.96
15	0.84	0.85	0.83	0.79	0.73	0.68	0.84	0.87	0.90	0.92	0.93	0.95
25	0.88	0.86	0.82	0.75	0.70	0.65	0.95	0.95	0.97	0.97	1.00	1.03
50	0.93	0.94	0.95	0.97	0.86	0.73	0.98	0.97	0.95	0.91	0.78	0.72
			$\rho=0.7,\phi=0.5$						$\rho=0.3,\phi=0.5$			
10	0.95	0.99	1.04	1.14	1.23	1.33	0.87	0.91	0.94	0.97	0.99	1.00
15	0.92	0.97	1.03	1.07	1.05	1.04	0.84	0.87	0.91	0.93	0.94	0.95
25	0.97	0.98	0.99	0.99	1.00	1.01	0.93	0.94	0.94	0.91	0.89	0.86
50	0.99	0.99	0.98	0.97	0.92	0.88	0.96	0.97	0.96	0.95	0.85	0.79

Table entries are the ratio of the average QFL selected by the panel FPE minimization (15) to the average QFL selected by the conventional FPE minimization (26). Fixed effects case ($p=1$). Entries less than unity indicate the panel FPE criterion selects a lower QFL on average.

TABLE A10. Comparison of panel FPE forecast selection to KLIC forecast selection

T	$n=10$	$n=15$	$n=25$	$n=50$	$n=100$	$n=200$	$n=10$	$n=15$	$n=25$	$n=50$	$n=100$	$n=200$
			$\rho=-0.7,\phi=0$						$\rho=-0.3,\phi=0$			
10	0.98	0.94	0.86	0.77	0.70	0.64	1.01	0.97	0.94	0.92	0.94	0.98
15	1.02	0.93	0.84	0.74	0.66	0.60	1.03	0.97	0.93	0.92	0.94	0.98
25	0.99	0.91	0.81	0.75	0.68	0.65	0.99	0.95	0.93	0.92	0.93	0.95
50	1.00	0.99	0.98	1.00	0.85	0.72	1.00	0.96	0.91	0.86	0.68	0.67
			$\rho=0.7,\phi=0$						$\rho=0.3,\phi=0$			
10	0.99	1.01	1.05	1.13	1.24	1.35	0.99	0.98	0.97	0.96	0.97	0.99
15	0.98	0.97	1.01	1.03	1.01	1.01	0.99	0.95	0.92	0.90	0.90	0.92
25	0.99	0.97	0.97	0.96	0.96	0.97	0.99	0.95	0.91	0.88	0.84	0.81
50	0.99	0.97	0.95	0.92	0.81	0.79	0.99	0.95	0.91	0.86	0.72	0.70
			$\rho=-0.7,\phi=0.5$						$\rho=-0.3,\phi=0.5$			
10	0.97	0.92	0.85	0.81	0.74	0.67	0.98	0.94	0.92	0.90	0.91	0.94
15	0.95	0.88	0.81	0.75	0.68	0.63	0.97	0.91	0.88	0.88	0.90	0.95
25	0.81	0.76	0.71	0.68	0.64	0.61	0.90	0.85	0.85	0.87	0.88	0.91
50	0.76	0.77	0.79	0.86	0.68	0.64	0.79	0.77	0.76	0.75	0.62	0.62
			$\rho=0.7,\phi=0.5$						$\rho=0.3,\phi=0.5$			
10	0.99	1.00	1.03	1.09	1.19	1.28	0.98	0.95	0.96	0.95	0.96	0.98
15	0.96	0.96	1.00	1.02	1.02	1.00	0.95	0.91	0.89	0.87	0.88	0.90
25	0.97	0.95	0.96	0.92	0.93	0.95	0.89	0.86	0.85	0.82	0.80	0.80
50	0.90	0.90	0.89	0.89	0.78	0.76	0.78	0.77	0.78	0.78	0.68	0.68

Table entries are the ratio of the average QFL selected by the panel FPE minimization (15) to the average QFL selected by KLIC (27). Fixed effects case ($p=1$). Entries less than unity indicate the panel FPE criterion selects a lower QFL on average.

TABLE A11. Comparison of LS panel FPE forecast selection to BCLS panel FPE forecast selection

T	$n=10$	$n=15$	$n=25$	$n=50$	$n=100$	$n=200$	$n=10$	$n=15$	$n=25$	$n=50$	$n=100$	$n=200$
	$\rho=-0.7,\phi=0$						$\rho=-0.3,\phi=0$					
10	1.41	1.45	1.45	1.42	1.37	1.31	1.37	1.56	1.58	1.55	1.55	1.55
15	1.35	1.38	1.38	1.33	1.27	1.21	1.37	1.43	1.43	1.39	1.37	1.37
25	1.18	1.16	1.13	1.18	1.11	1.08	1.26	1.29	1.29	1.28	1.24	1.21
50	1.04	1.04	1.04	1.05	1.05	1.04	1.13	1.16	1.18	1.18	1.18	1.18
	$\rho=0.7,\phi=0$						$\rho=0.3,\phi=0$					
10	1.09	1.19	1.30	1.36	1.40	1.41	1.20	1.34	1.47	1.57	1.65	1.69
15	1.15	1.10	1.30	1.40	1.48	1.52	1.29	1.40	1.49	1.55	1.57	1.58
25	1.25	1.31	1.39	1.45	1.51	1.54	1.33	1.39	1.42	1.44	1.44	1.41
50	1.18	1.23	1.28	1.33	1.38	1.39	1.18	1.21	1.23	1.26	1.28	1.30
	$\rho=-0.7,\phi=0.5$						$\rho=-0.3,\phi=0.5$					
10	1.33	1.41	1.45	1.45	1.40	1.35	1.34	1.48	1.59	1.57	1.56	1.55
15	1.28	1.34	1.39	1.37	1.31	1.25	1.26	1.34	1.43	1.42	1.39	1.37
25	1.17	1.18	1.16	1.21	1.15	1.10	1.23	1.26	1.29	1.30	1.26	1.23
50	1.03	1.04	1.04	1.05	1.04	1.04	1.09	1.14	1.17	1.19	1.19	1.22
	$\rho=0.7,\phi=0.5$						$\rho=0.3,\phi=0.5$					
10	0.97	1.09	1.20	1.31	1.37	1.40	1.07	1.20	1.35	1.50	1.61	1.66
15	0.84	1.10	1.22	1.34	1.42	1.49	1.11	1.25	1.41	1.51	1.56	1.58
25	1.15	1.22	1.31	1.39	1.46	1.52	1.24	1.32	1.39	1.42	1.44	1.43
50	1.13	1.19	1.23	1.30	1.36	1.40	1.13	1.18	1.21	1.25	1.28	1.30

Table entries are the ratio of the average QFL of the LS forecast selected by panel FPE minimization (15) to the average QFL of the BCLS forecast selected by BCLS panel FPE minimization (23). Fixed effects case ($p = 1$). Entries less than unity indicate a lower LS QFL on average.

TABLE A12. Comparison of LS panel FPE forecast selection to BCLS panel FPE forecast selection

T	$n=10$	$n=15$	$n=25$	$n=50$	$n=100$	$n=200$	$n=10$	$n=15$	$n=25$	$n=50$	$n=100$	$n=200$
			$\rho=-0.7,\phi=0$						$\rho=-0.3,\phi=0$			
10	0.02	0.05	0.06	0.10	0.10	0.10	0.03	0.05	0.05	0.04	0.02	0.00
15	0.05	0.12	0.16	0.20	0.19	0.18	0.08	0.08	0.12	0.10	0.04	0.00
25	0.21	0.27	0.34	0.47	0.44	0.19	0.25	0.34	0.51	0.46	0.39	0.18
50	0.17	0.24	0.28	0.24	0.29	0.30	0.25	0.33	0.49	0.81	1.89	3.09
			$\rho=0.7,\phi=0$						$\rho=0.3,\phi=0$			
10	−0.01	−0.02	−0.02	−0.03	−0.03	−0.03	0.06	0.05	0.05	0.02	0.01	0.00
15	0.03	0.04	0.09	0.10	0.08	0.05	0.06	0.10	0.12	0.08	0.04	0.02
25	0.16	0.22	0.28	0.33	0.37	0.33	0.22	0.26	0.37	0.40	0.46	0.48
50	0.13	0.19	0.38	0.51	1.18	1.34	0.18	0.30	0.51	0.86	2.06	2.83
			$\rho=-0.7,\phi=0.5$						$\rho=-0.3,\phi=0.5$			
10	0.00	0.03	0.04	0.08	0.11	0.09	0.01	0.03	0.03	0.04	0.03	0.01
15	0.00	0.06	0.10	0.17	0.21	0.21	0.03	0.07	0.10	0.12	0.07	0.02
25	0.14	0.18	0.31	0.44	0.47	0.34	0.16	0.24	0.35	0.38	0.40	0.35
50	0.15	0.20	0.23	0.25	0.49	0.73	0.18	0.24	0.33	0.48	1.63	2.44
			$\rho=0.7,\phi=0.5$						$\rho=0.3,\phi=0.5$			
10	−0.01	−0.02	−0.02	−0.03	−0.03	−0.03	0.02	0.04	0.06	0.04	0.02	0.00
15	0.03	0.04	0.09	0.10	0.08	0.05	0.04	0.08	0.08	0.12	0.08	0.04
25	0.16	0.22	0.28	0.33	0.37	0.33	0.16	0.21	0.30	0.34	0.45	0.50
50	0.13	0.19	0.38	0.51	1.18	1.34	0.13	0.18	0.31	0.60	1.66	2.24

Table entries are the difference between the average number of lags selected by BCLS panel FPE minimization (23) and panel FPE minimization (15). Fixed effects case ($p=1$). Positive entries indicate that BCLS panel FPE minimization selected a larger lag order on average.

APPENDIX B: Proofs

B.1. Preliminary Definitions and Auxiliary Lemmas

Some preliminary results and simplifying notation are required in order to prove the Theorems of the article. Let $\delta_{n,T-k} := \min\left(n^{\frac{1}{2}}(T-k)^{\frac{1}{2}}, T-k\right)$, such that $k\delta_{n,T-k}^{-2} \le k_{n,T}\delta_{n,T-k}^{-2} = o(1)$ as $n \to \infty$ and $T \to \infty$. C_q will denote a finite positive constant that is dependent on an arbitrary positive integer q. It takes on different values at different places. We also define

$$l(k) := \mathrm{tr}\left(\Phi L_{n,T}(k)\right), \quad \|\boldsymbol{\alpha}(k) - \boldsymbol{\alpha}\|^2_{\Gamma\otimes\Phi} := \mathrm{tr}(\Phi\Lambda(k)),$$

and we use $\|\mathbf{A}\|_{\mathbf{B}}$ to denote $\|\mathbf{A}\|_{\mathbf{B}} = \mathrm{tr}^{\frac{1}{2}}\left(\mathbf{A}'\mathbf{B}\mathbf{A}\right)$ for conformable matrices $\mathbf{A}$ and $\mathbf{B}$.

Time Trend Polynomial. Let $\boldsymbol{\eta}_N$ be a $N \times p$ matrix defined such that $\boldsymbol{\eta}_N'\boldsymbol{\eta}_N = I_p$ and $\boldsymbol{\eta}_N\left(\boldsymbol{\eta}_N'\boldsymbol{\eta}_N\right)^{-1}\boldsymbol{\eta}_N'\varsigma_N := \varsigma_N$, so that $\boldsymbol{\eta}_N$ is an orthonormal basis of the time trend polynomials ς_N. Let the set of $p \times 1$ vectors $\left\{\eta_t^{(N)}\right\}_{t=1}^{N}$ be defined such that $\boldsymbol{\eta}_N' := \left[\eta_1^{(N)} : \cdots : \eta_N^{(N)}\right]$, and let the scalars $\left\{\eta_{q,t}^{(N)}\right\}_{q=1}^{p}$ be defined such that $\eta_t^{(N)} = \left(\eta_{1,t}^{(N)}, \ldots, \eta_{p,t}^{(N)}\right)'$. Using Corollary 3.1 of Eisinberg and Fedele (2007) we can solve for

$$\eta_{q,t}^{(N)} := \left(\frac{2q-1}{N}\frac{\prod_{r=0}^{q-1}(N-r)}{\prod_{r=0}^{q-1}(N+r)}\right)^{\frac{1}{2}} d_{q,t}^{(N)}, \text{ for } t = 1, 2, \ldots, N, N+1, \tag{B.1}$$

where $d_{1,t}^{(N)} = 1$ for all $t = 1, 2, \ldots$ and

$$d_{q,t}^{(N)} := \sum_{s=1}^{q}(-1)^{s+q}\left(\frac{\prod_{r=1}^{q-1}(s+r-1)}{\prod_{r=1}^{q-s} r}\cdot\frac{\prod_{r=1}^{s-1}(t-s+r)}{\prod_{r=1}^{s-1}(N-r)}\cdot\frac{1}{\prod_{r=1}^{s-1} r}\right), \text{ for } q = 2, 3, \ldots, p, \tag{B.2}$$

so that $d_{q,t}^{(N)}$ satisfies $\left|d_{q,t}^{(N)}\right| \le 1$ for t and q satisfying $1 \le t \le N$ and $1 \le q \le p$, respectively.

Prediction Error. Let $\ddot{X}_{i,t}(k,l) := X_{i,t}(k) - \bar{X}_{i,t}(k,l)$, where $\bar{X}_{i,t}(k,l) := \mathbf{X}_i^{(k,l)\prime}\boldsymbol{\eta}_{T-l}\eta_{t-l+1}^{(T-l)}$ and $\mathbf{X}_i^{(k,l)} := \left[X_{i,l}(k) : \cdots : X_{i,T-1}(k)\right]'$. Then, using a similar approach to Ing (2003, p. 276), we can re-express $\hat{y}_{i,t+1}(k,l) = \hat{\boldsymbol{\alpha}}(k,l)'\ddot{X}_{i,t}(k,l) + \mathbf{y}_i^{(l)\prime}\boldsymbol{\eta}_{T-l}\eta_{t-l+1}^{(T-l)}$. Using this expression, for each $k = 1, 2, \ldots$ we can decompose

$$\hat{u}_{i,t+1}(k,l) = \tfrac{1}{n(T-l)}\sum_{j=1}^{n}\mathbf{e}_j^{(l)\prime}\mathbf{M}_{T-l}\mathbf{X}_j^{(k,l)}\hat{\mathbf{Q}}_{(k,l)}^{-1}\ddot{X}_{i,t}(k,l) - \ddot{e}_{i,t+1}(k,l) +$$
$$\tfrac{1}{n(T-l)}\sum_{j=1}^{n}\mathbf{s}_j^{(k,l)\prime}\mathbf{M}_{T-l}\mathbf{X}_j^{(k,l)}\hat{\mathbf{Q}}_{(k,l)}^{-1}\ddot{X}_{i,t}(k,l) - \ddot{s}_{i,t}(k,l) =: \sum_{r=1}^{4}\varepsilon_{r,i,t+1}^{(k,l)}, \tag{B.3}$$

say, where $\ddot{e}_{i,t+1}(l) := e_{i,t+1} - \bar{e}_{i,t+1}(l)$, $\bar{e}_{i,t+1}(l) := \mathbf{e}_i^{(l)\prime}\boldsymbol{\eta}_{T-l}\eta_{t-l+1}^{(T-l)}$, and $\mathbf{e}_i^{(l)} := \left[e_{i,l+1} : \cdots : e_{i,T}\right]'$; and $\ddot{s}_{i,t}(k,l) := s_{i,t}(k) - \bar{s}_{i,t}(k,l)$, $\bar{s}_{i,t}(k,l) := \mathbf{s}_i^{(k,l)\prime}\boldsymbol{\eta}_{T-l}\eta_{t-l+1}^{(T-l)}$, and $\mathbf{s}_i^{(k,l)} := \left[s_{i,l}(k) : \cdots : s_{i,T-1}(k)\right]'$. We also define $\mathbf{u}_i^{(k,l)} := \mathbf{e}_i^{(l)} + \mathbf{s}_i^{(k,l)}$.

B.1.1. Auxiliary Lemmas and Theorems. We make extensive use of the Theorems below, which are generalizations of the First Moment Bound Theorem of Findley and Wei (1993) to a panel data setting.

THEOREM B.1. *Notation defined herein only applies within this Theorem. Consider a linear vector moving average process of the form $y_{i,t} = \sum_{s=0}^{\infty} \theta'_{s,i} e_{i,t-s}$, satisfying the following properties*:

(i) *For each $i = 1,2,\ldots$, the sequence of matrices $\{\theta_{s,i}\}_{s=0}^{\infty}$ satisfies $\sum_{s=0}^{\infty} \|\theta_{s,i}\| < \infty$.*

(ii) *For each $i = 1,2,\ldots$, the error vector $e_{i,t} \sim iid(0,\Sigma_i)$, $\Sigma_i > 0$, and $\sup_{-\infty<t<\infty} E\|e_{i,t}\|^{8q} < \infty$.*

Let $x_{i,t}$ be another linear vector process satisfying properties (i) *and* (ii). *Then for any nonrandom scalar coefficients $h_{i,j}$ and $k_{s,t}$* (*permitted to be dependent on n and T, respectively*),

$$E\left\| \sum_{i,j=1}^{n} h_{i,j} \sum_{s,t=1}^{T} k_{s,t} \left(x_{i,t} y'_{j,s} - E\left(x_{i,t} y'_{j,s} \right) \right) \right\|^{4q} \leq C_q \left(\sum_{i,j,k,l=1}^{n} h_{i,j} h_{k,l} \sum_{s,t,u,v=1}^{T} k_{s,t} k_{u,v} E\left(x'_{i,t} x_{l,v} \right) E\left(y'_{j,s} y_{k,u} \right) \right)^{2q},$$

where C_q is a finite constant dependent only on q.

Proof. The proof follows the same arguments as given in the proof to Theorem A.1 in Greenaway-McGrevy (2015). ■

THEOREM B.2. *Let $w_{i,t}$ and $z_{i,t}$ be vector moving average processes satisfying properties* (i) *and* (ii) *of Theorem B.1. Then for any nonrandom coefficients $g_{i,j}$ and $h_{i,j}$* (*permitted to be dependent on n*), *and $l_{s,t}$ and $k_{s,t}$* (*permitted to be dependent T*),

$$E\left\| \sum_{i,j=1}^{n} g_{i,j} h_{i,j} \sum_{s,t=1}^{T} k_{s,t} l_{s,t} \left(\left(x_{i,t} y'_{j,s} - E\left(x_{i,t} y'_{j,s} \right) \right) \otimes \left(w_{i,t} z'_{j,s} - E\left(w_{i,t} z'_{j,s} \right) \right) \right) \right\|^{2q}$$
$$\leq C_q \left(\sum_{i,j,k,l=1}^{n} h_{i,j} h_{k,l} g_{i,j} g_{k,l} \sum_{s,t,u,v=1}^{T} k_{s,t} k_{u,v} l_{s,t} l_{u,v} E(x'_{i,t} x_{l,v}) E(w'_{i,t} w_{l,v}) E\left(y'_{j,s} y_{k,u} \right) E\left(z'_{j,s} z_{k,u} \right) \right)^{q},$$

where $y_{i,t}$ and $x_{i,t}$ are the same processes defined in Theorem B.1.

Proof. The proof follows the same arguments as given in the proof to Theorem A.1 in Greenaway-McGrevy (2015). ■

We also make extensive use of the following Lemma.

LEMMA B.3. *Under Assumptions 1 and 2, for all $q = 1,2,\ldots$ and integers k and l satisfying $1 \leq k \leq l \leq k_{n,T}$,*

(i) $\mathrm{E}\left\| \frac{1}{n(T-l)} \sum_{i=1}^{n} \mathbf{X}_i^{(k,l)\prime} \mathbf{e}_i^{(l)} \right\|^q \leq C_q \left(\left(\frac{k}{n(T-l)} \right)^{\frac{q}{2}} \right)$

(ii) $\mathrm{E}\left\| \frac{1}{n(T-l)} \sum_{i=1}^{n} \mathbf{X}_i^{(k,l)\prime} \mathbf{P}_{T-l} \mathbf{e}_i^{(l)} - \frac{p}{T-l} \boldsymbol{\xi}(k) \right\|^q \leq C_q \left(\frac{k}{n(T-l)^2} \right)^{\frac{q}{2}}$

$$(iii)\quad \mathrm{E}\left\|\tfrac{1}{n(T-l)}\sum_{i=1}^{n}\mathbf{X}_i^{(k,l)\prime}\mathbf{M}_{T-l}\mathbf{e}_i^{(l)}\right\|^q \le C_q\left(\left(\tfrac{k}{n(T-l)}\right)^{\frac{q}{2}}+\left(\tfrac{k}{(T-l)^2}\right)^{\frac{q}{2}}\right)$$

$$(iv)\quad \mathrm{E}\left\|\tfrac{1}{n}\sum_{i=1}^{n}\left(\tfrac{1}{T-l}\mathbf{e}_i^{(l)\prime}\mathbf{e}_i^{(l)}-\Sigma_i\right)\right\|^q \le C_q n^{-\frac{q}{2}}(T-l)^{-\frac{q}{2}}$$

$$(v)\quad \mathrm{E}\left\|\tfrac{1}{n}\sum_{i=1}^{n}\left(\mathbf{e}_i^{(l)\prime}\mathbf{P}_{T-l}\mathbf{e}_i^{(l)}-p\Sigma_i\right)\right\|^q \le C_q n^{-\frac{q}{2}}(T-l)^{-\frac{q}{2}}$$

$$(vi)\quad \mathrm{E}\left\|\tfrac{1}{n(T-l)}\sum_{i=1}^{n}\mathbf{s}_i^{(k,l)\prime}\mathbf{s}_i^{(k,l)}-\Lambda(k)\right\|^q \le C_q n^{-\frac{q}{2}}(T-l)^{-\frac{q}{2}}\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{2q}$$

$$(vii)\quad \mathrm{E}\left\|\tfrac{1}{n(T-l)}\sum_{i=1}^{n}\mathbf{s}_i^{(k,l)\prime}\mathbf{P}_{T-l}\mathbf{s}_i^{(k,l)}\right\|_{\Phi}^q \le C_q(T-l)^{-q}\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{2q}$$

$$(viii)\quad \mathrm{E}\left\|\tfrac{1}{n(T-l)}\sum_{i=1}^{n}\mathbf{X}_i^{(k,l)\prime}\mathbf{s}_i^{(k,l)}\right\|_{\Phi}^q \le C_q\left(\left(\tfrac{k}{n(T-l)}\right)^{\frac{q}{2}}\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{q}\right)$$

$$(ix)\quad \mathrm{E}\left\|\tfrac{1}{n(T-l)}\sum_{i=1}^{n}\mathbf{X}_i^{(k,l)\prime}\mathbf{M}_{T-l}\mathbf{s}_i^{(k,l)}\right\|_{\Phi}^q \le C_q\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{q}\left(\left(\tfrac{k}{n(T-l)}\right)^{\frac{q}{2}}+\left(\tfrac{k}{(T-l)^2}\right)^{\frac{q}{2}}\right)$$

$$(x)\quad \mathrm{E}\left\|\tfrac{1}{n(T-l)}\sum_{i=1}^{n}\mathbf{e}_i^{(l)\prime}\mathbf{M}_{T-l}\mathbf{s}_i^{(k,l)}\right\|_{\Phi}^q \le C_q\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{q}\left((n(T-l))^{-\frac{q}{2}}+(T-l)^{-q}\right)$$

$$(xi)\quad \mathrm{E}\left\|\tfrac{1}{n(T-l)}\sum_{i=1}^{n}\mathbf{X}_i^{(k,l)\prime}\mathbf{M}_{T-l}\mathbf{X}_i^{(k,l)}-\Gamma(k)\right\|^q \le C_q\left(\left(\tfrac{k}{n(T-l)}\right)^{\frac{q}{2}}+\left(\tfrac{k}{(T-l)^2}\right)^{\frac{q}{2}}\right)$$

$$(xii)\quad \mathrm{E}\left\|\hat{\mathbf{Q}}_{(k,l)}^{-1}\right\|^q = O(1).$$

Proof. Lemmas (i) through (xi) follow by straightforward application of Theorem B.1. We provide an example of how to apply the Theorem to prove (ii). First note that by the definition of $\boldsymbol{\xi}(k)$, $\left\|\mathrm{E}\left(\frac{1}{n}\sum_i \mathbf{X}_i^{(k,l)\prime}\mathbf{P}_{T-l}\mathbf{e}_i^{(l)}\right)-p\boldsymbol{\xi}(k)\right\|^q \le C_q\left(k^{\frac{q}{2}}(T-l)^{-q}\right)$. Then by Theorem B.1 we have

$$\mathrm{E}\left\|\tfrac{1}{n(T-l)}\sum_{i=1}^{n}\mathbf{X}_i^{(k,l)\prime}\mathbf{P}_{T-l}\mathbf{e}_i^{(l)}-\tfrac{p}{T-l}\boldsymbol{\xi}(k)\right\|^q$$
$$\le C_q\left\|\tfrac{1}{n^2(T-l)^2}\sum_{i,j=1}^{n}\mathrm{E}\left(\mathbf{X}_i^{(k,l)\prime}\mathbf{P}_{T-l}\mathbf{X}_j^{(k,l)}\right)\mathrm{E}\left(\mathbf{e}_i^{(l)\prime}\mathbf{P}_{T-l}\mathbf{e}_j^{(l)}\right)\right\|^{\frac{q}{2}}+C_q\left(\tfrac{k}{(T-l)^4}\right)^{\frac{q}{2}}.$$

The stated result then follows since under (B.1) and (B.2),

$$\tfrac{1}{n^2(T-l)^2}\sum_{i,j=1}^{n}\mathrm{E}\left(\mathbf{X}_i^{(k,l)\prime}\mathbf{P}_{T-l}\mathbf{X}_j^{(k,l)}\right)\mathrm{E}\left(\mathbf{e}_i^{(l)\prime}\mathbf{P}_{T-l}\mathbf{e}_j^{(l)}\right)$$
$$\le \tfrac{k}{n^2(T-l)^4}\sum_{i,j=1}^{n}\sum_{s,t,u,v=k+1}^{T-1}\mathrm{E}\left(X_{i,t}'(k)X_{j,s}(k)\right)\mathrm{E}\left(e_{i,u+1}'e_{j,v+1}\right)\le \tfrac{kp^2}{n(T-l)^2}\|C_\Sigma\|^2\|C_\theta\|^2.$$

To show (xii) we follow similar steps to the proof of Theorem 2 (ii) of Ing and Wei (2003). First, under the Lipschitz condition we can straightforwardly establish that for any $\varepsilon>0$ we have

$$\mathrm{E}\left\|\hat{\mathbf{Q}}_{(k,l)}^{-1}\right\|^q \le C_q k^{(2+\varepsilon)q}. \qquad \textbf{(B.4)}$$

Next, under Lemma (xii),

$$\mathrm{E}\left\|\hat{\mathbf{Q}}_{(k,l)}-\Gamma(k)\right\|^q \le C_q k^q \delta_{n,T-l}^{-q}. \qquad \textbf{(B.5)}$$

Then by recursive substitution of (B.4) and (B.5) into the following inequality s times,

$$\mathrm{E}\left\|\hat{\mathbf{Q}}_{(k,l)}^{-1}\right\|^{q} \leq C_q\left(1+\left(\mathrm{E}\left\|\hat{\mathbf{Q}}_{(k,l)}-\Gamma(k)\right\|^{2q}\right)^{\frac{1}{2}}\left(\mathrm{E}\left\|\hat{\mathbf{Q}}_{(k,l)}^{-1}\right\|^{2q}\right)^{\frac{1}{2}}\right)$$

we can obtain $\mathrm{E}\left\|\hat{\mathbf{Q}}_{(k,l)}^{-1}\right\|^{q} \leq C_q\left(1+k^{(2+\varepsilon)q}\left(\delta_{n,T-l}^{-1}k\right)^{sq}\right)$. Thus we can choose an arbitrarily large s to ensure that $k^{(2+\varepsilon)q}\left(\delta_{n,T-l}^{-1}k\right)^{sq}=o(1)$. ■

B.1.2. Bounds and Limits. We make extensive use of the following bound and limits.

$$\max_{1\leq k\leq k_{n,T}}\left(l^{-1}(k)\,k\delta_{n,T-k}^{-2}\right)\leq C, \tag{B.6}$$

$$\max_{1\leq k\leq k_{n,T}}\left(l^{-1}(k)\left\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{2}\right)\leq 1. \tag{B.7}$$

The following is a consequence of (B.7) and $k_{n,T}\delta_{n,T-k}^{-2}=o(1)$.

$$\lim_{n,T\to\infty}\max_{1\leq k\leq k_{n,T}}\left(l^{-1}(k)\,k\delta_{n,T-k}^{-2}\left\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{2}\right)=0. \tag{B.8}$$

Finally

$$\lim_{n,T\to\infty}\max_{1\leq k\leq k_{n,T}}\left(l^{-1}(k)\,\delta_{n,T-k}^{-2}\right)=0, \tag{B.9}$$

which follows once we establish that $l^{-1}(k)\,\delta_{n,T-k}^{-2}\leq C^{-1}\left(k_{n,T}^{*}\right)^{-1}$ for some finite $C>0$, since $k_{n,T}^{*}\to\infty$. First note that $\delta_{n,T-k}^{2}l(k)\geq Ck$, so that for all $k>k_{n,T}^{*}$, $l^{-1}(k)\,\delta_{n,T-k}^{-2}\leq C^{-1}k^{-1}<C^{-1}\left(k_{n,T}^{*}\right)^{-1}$. Next, for $k\leq k_{n,T}^{*}$, since $l(k)\geq l\left(k_{n,T}^{*}\right)$, we have $l^{-1}(k)\,\delta_{n,T-k}^{-2}\leq l^{-1}\left(k_{n,T}^{*}\right)\delta_{n,T-k}^{-2}\leq C^{-1}\left(k_{n,T}^{*}\right)^{-1}$.

B.2. Proofs of Least Squares Theorems

Sections B.2.1 and B.2.3 below provide the proofs to Theorems 3.1 and 3.2, respectively. In Section B.2.2 we establish that the QFR estimator is asymptotically unbiased.

B.2.1. Proof of Theorem 3.1 (*Asymptotic QFR*). The proof follows the same strategy as the proof to Theorem 3.1 of Greenaway-McGrevy (2015), allowing for cross sectional heterogeneity and weak dependence. We also solve for the asymptotic QFR when $k_{n,T}T^{-1/2}=o(1)$, whereas Greenaway-McGrevy (2015) only considers $k_{n,T}n^{-1/2}=o(1)$. The out-of sample forecast error is given by setting $t=T$ and $l=k$ in (B.3). We can then decompose

$$\mathcal{L}(k)=\tfrac{1}{n}\sum\nolimits_{i=1}^{n}\hat{u}_{i,T+1}(k,k)\,\hat{u}_{i,T+1}'(k,k)=\sum\nolimits_{r,r'=1}^{4}H_{(r,r')}(k),$$

where
$H_{(r,r')}(k) := \frac{1}{n}\sum_{i=1}^{n} \varepsilon_{r,i,T+1}^{(k,k)}(k)\,\varepsilon_{r',i,T+1}^{(k,k)\prime}$, $(r,r') \in \{1,\ldots,4\}$,

for $\left\{\varepsilon_{r,i,T+1}^{(k,k)}\right\}_{r=1}^{4}$ defined in (B.3). Using the ten unique terms comprising this sum, the proof derives upper bounds for $l^{-1}(k) \cdot \left|\mathrm{tr}\left(\Phi\left(\mathrm{E}(\mathcal{L}(k)) - \Sigma\left(1+p^2T^{-1}\right) - L_{n,T}(k)\right)\right)\right|$, and then uses (B.6) through (B.9) to show that the bounds are $o(1)$ for all k as $n \to \infty$ and $T \to \infty$.

For brevity, throughout the remainder of this subsection we omit the k signifiers, so that $\hat{\mathbf{Q}} := \hat{\mathbf{Q}}_{(k,k)}$, $\mathbf{X}_i := \mathbf{X}_i^{(k,k)}$, $\mathbf{u}_i := \mathbf{u}_i^{(k,k)}$, $\mathbf{s}_i := \mathbf{s}_i^{(k,k)}$, $\mathbf{e}_i := \mathbf{e}_i^{(k)}$, $\mathbf{M} := \mathbf{M}_{T-k}$, $\mathbf{P} := \mathbf{P}_{T-k}$, $\eta_t := \eta_t^{(T-k)}$, $\boldsymbol{\eta} := \boldsymbol{\eta}_{T-k}$, $\mathbf{o} := \mathbf{o}(k)$ and $\boldsymbol{\xi} := \boldsymbol{\xi}(k)$. Similarly, $\bar{e}_{i,t+1} := \mathbf{e}_i'\boldsymbol{\eta}\eta_{t-k+1}$, $\bar{X}_{i,t} := \mathbf{X}_i'\boldsymbol{\eta}\eta_{t-k+1}$ and $\bar{s}_{i,t}(k) := \mathbf{s}_i'\boldsymbol{\eta}\eta_{t-k+1}$; and $\ddot{e}_{i,t+1} := e_{i,t+1} - \bar{e}_{i,t+1}$, $\ddot{X}_{i,t} := X_{i,t} - \bar{X}_{i,t}$ and $\ddot{s}_{i,t}(k) := s_{i,t} - \bar{s}_{i,t}(k)$.

For $H_{(1,1)}$ we first establish the following result. Let $\ddot{\mathbf{Q}}_T := \frac{1}{n}\sum_i \ddot{X}_{i,T}(k)\,\ddot{X}_{i,T}'(k)$ and $\hat{\mathbf{Q}}_T := \frac{1}{n}\sum_i X_{i,T}(k)\,X_{i,T}'(k)$. By Theorem B.1, Hölder's inequality, and Lemmas B.3(xi) and (xii),

$$\mathrm{E}\left\|\hat{\mathbf{Q}}^{-1}\ddot{\mathbf{Q}}_T\hat{\mathbf{Q}}^{-1} - \Gamma^{-1}\hat{\mathbf{Q}}_T\Gamma^{-1}\right\|^q \le C_q\left(\frac{k^q}{n^{q/2}(T-k)^{q/2}} + \frac{k^q}{(T-k)^q}\right),\ q = 1,2,\ldots, \tag{B.10}$$

where $k(T-k)^{-1} = o(1)$. By Hölder's inequality and Lemma B.3(iii) this means that

$$H_{(1,1)} = \frac{1}{n^2(T-k)^2}\sum_{i,j=1}^{n} \mathbf{e}_i'\mathbf{M}\mathbf{X}_i\Gamma^{-1}\hat{\mathbf{Q}}_T\Gamma^{-1}\mathbf{X}_j'\mathbf{M}\mathbf{e}_j + o_{n,T}^{(a)}(k),$$

where $o_{n,T}^{(a)}(k)$ denotes a negligible term satisfying $\mathrm{E}\left\|o_{n,T}^{(a)}(k)\right\| = o\left(k\delta_{n,T-k}^{-2}\right)$. Using Lemma B.3(ii) and Hölder's inequality, the first term in the expression above satisfies

$$\begin{aligned}&\frac{1}{n^2(T-k)^2}\sum_{i,j=1}^{n} \mathbf{e}_i'\mathbf{M}\mathbf{X}_i\Gamma^{-1}\hat{\mathbf{Q}}_T\Gamma^{-1}\mathbf{X}_j'\mathbf{M}\mathbf{e}_j\\ &= \frac{1}{n^2(T-k)^2}\sum_{i,j=1}^{n} \mathbf{e}_i'\mathbf{X}_i\Gamma^{-1}\hat{\mathbf{Q}}_T\Gamma^{-1}\mathbf{X}_j'\mathbf{e}_j - \frac{1}{n(T-k)^2}\sum_{i=1}^{n} \mathbf{e}_i'\mathbf{X}_i\Gamma^{-1}\hat{\mathbf{Q}}_T\Gamma^{-1}\boldsymbol{\xi} -\\ &\quad \frac{1}{n(T-k)^2}\sum_{j=1}^{n} \boldsymbol{\xi}'\Gamma^{-1}\hat{\mathbf{Q}}_T\Gamma^{-1}\mathbf{X}_j'\mathbf{e}_j + \frac{1}{(T-k)^2}\boldsymbol{\xi}'\Gamma^{-1}\hat{\mathbf{Q}}_T\Gamma^{-1}\boldsymbol{\xi} + o_{n,T}^{(b)}(k),\end{aligned}$$

where $o_{n,T}^{(b)}(k)$ denotes a negligible term satisfying $\mathrm{E}\left\|o_{n,T}^{(b)}(k)\right\| = o\left(k\delta_{n,T-k}^{-2}\right)$. Under Assumption 2(iii), $\mathrm{E}\left\|\hat{\mathbf{Q}}_T - \Gamma\right\|^{2q} = O\left(k^{2q}n^{-q}\right)$ for all $q = 1,2,\ldots$, and thus if $k_{n,T}n^{-1/2} = o(1)$ holds under Assumption 3, by application of Hölder's inequality and Lemma B.3(i) to the above expression it is straightforward to establish that

$$\begin{aligned}&\mathrm{tr}\left(\mathrm{E}\left(\frac{1}{n^2(T-k)^2}\sum_{i,j=1}^{n} \mathbf{e}_i'\mathbf{M}\mathbf{X}_i\Gamma^{-1}\hat{\mathbf{Q}}_T\Gamma^{-1}\mathbf{X}_j'\mathbf{M}\mathbf{e}_j\Phi\right)\right) - \frac{p^2}{(T-k)^2}\mathrm{tr}\left(\boldsymbol{\xi}'\Gamma^{-1}\boldsymbol{\xi}\Phi\right)\\ &\qquad - \frac{1}{n^2(T-k)}\sum_{i,j=1}^{n}\mathrm{tr}\left(\Sigma_{i,j}\Phi\right)\mathrm{tr}\left(\Gamma^{-1}\Gamma_{i,j}\right) = o\left(k\delta_{n,T-k}^{-2}\right),\end{aligned} \tag{B.11}$$

noting that

$$\begin{aligned}\mathrm{tr}\left(\mathrm{E}\left(\frac{1}{n^2(T-k)^2}\sum_{i,j=1}^{n} \mathbf{e}_i'\mathbf{X}_i\Gamma^{-1}\mathbf{X}_j'\mathbf{e}_j\Phi\right)\right) &= \mathrm{tr}\left[\left(\Gamma^{-1}\otimes\Phi\right)\frac{1}{n^2(T-k)^2}\sum_{i,j=1}^{n}\sum_{t=k}^{T-1}\mathrm{E}\left(X_{i,t}X_{j,t}'\otimes e_{i,t+1}e_{j,t+1}'\right)\right]\\ &= \frac{1}{n^2(T-k)}\sum_{i,j=1}^{n}\mathrm{tr}\left(\Phi\Sigma_{i,j}\right)\mathrm{tr}\left(\Gamma^{-1}\Gamma_{i,j}\right).\end{aligned}$$

Applying the bound in (B.6) to (B.11) we have

$$\max_{1\le k\le k_{n,T}}\left|l^{-1}(k)\,\mathrm{tr}\left(\Phi \mathrm{E}\left(H_{(1,1)}-\frac{k}{n^2(T-k)}\sum_{i,j=1}^{n}\Sigma_{i,j}\cdot\mathrm{tr}\left(\Gamma_{i,j}\Gamma^{-1}\right)-\frac{p^2}{(T-k)^2}\boldsymbol{\xi}'\Gamma^{-1}\boldsymbol{\xi}\right)\right)\right|=o(1). \qquad \textbf{(B.12)}$$

If $k_{n,T}T^{-1/2}=o(1)$ holds instead of $k_{n,T}n^{-1/2}=o(1)$ under Assumption 3, we can no longer ensure that $\mathrm{E}\left\|\hat{\mathbf{Q}}_T-\Gamma\right\|^{2q}=O\left(k^{2q}n^{-q}\right)=o(1)$, and we must use a different approach. We define

$$X^*_{i,T}:=\left(x^{*\prime}_{i,T},\ldots,x^{*\prime}_{i,T-k+1}\right),\;\; x^*_{i,T-r+1}:=\sum\nolimits_{s=0}^{N_0-k}\theta'_s e_{i,T-r-s+1},\; r=1,\ldots,k, \qquad \textbf{(B.13)}$$

where N_0 is the largest integer less than $\sqrt{T-k}$, and we assume $k<N_0$. Let $X^{**}_{i,T}:=X_{i,T}-X^*_{i,T}$, which satisfies

$$\frac{1}{n}\sum\nolimits_{i=1}^{n}\mathrm{E}\left\|X^{**}_{i,T}\right\|^2\le C\frac{k}{\sqrt{T-k}}\left(\sum\nolimits_{s=N_0+1}^{\infty}s\left\|\theta_s\right\|^2\right)=O\left(k(T-k)^{-\frac{1}{2}}\right),$$

which is $o(1)$ when $kT^{-1/2}=o(1)$. Using this result and decomposing

$$\hat{\mathbf{Q}}_T=\frac{1}{n}\sum\nolimits_{i=1}^{n}X^{**}_{i,T}X^{**\prime}_{i,T}+\frac{1}{n}\sum\nolimits_{i=1}^{n}X^{*}_{i,T}X^{**\prime}_{i,T}+\frac{1}{n}\sum\nolimits_{i=1}^{n}X^{**}_{i,T}X^{*\prime}_{i,T}+\frac{1}{n}\sum\nolimits_{i=1}^{n}X^{*}_{i,T}X^{*\prime}_{i,T}, \qquad \textbf{(B.14)}$$

we can follow the same method as used in the proof to Theorem 3 of Ing and Wei (2003) to establish

$$\frac{1}{n^2(T-k)^2}\sum_{i,j=1}^{n}\mathbf{e}'_i\mathbf{X}_i\Gamma^{-1}\hat{\mathbf{Q}}_T\Gamma^{-1}\mathbf{X}'_j\mathbf{e}_j=\frac{1}{n^2(T-k)^2}\sum_{s,t=k}^{T-N_0-1}\sum_{i,j,h=1}^{n}e_{i,t+1}X'_{i,t}\Gamma^{-1}X^*_{j,T}X^{*\prime}_{j,T}\Gamma^{-1}X_{h,s}e_{h,s+1}+o^{(c)}_{n,T}(k),$$

where $o^{(c)}_{n,T}(k)$ denotes a negligible term satisfying $\mathrm{E}\left\|o^{(c)}_{n,T}(k)\right\|=o\left(k\delta^{-2}_{n,T-k}\right)$. Now

$$\mathrm{tr}\left(\mathrm{E}\left(\frac{1}{n^2(T-k)^2}\left(\sum\nolimits_{t=k}^{T-N_0-1}\sum\nolimits_{i=1}^{n}e_{i,t+1}X_{i,t}\right)\Gamma^{-1}\left(\frac{1}{n}\sum\nolimits_{i=1}^{n}X^*_{i,T}X^{*\prime}_{i,T}\right)\Gamma^{-1}\left(\sum\nolimits_{t=k}^{T-N_0-1}\sum\nolimits_{i=1}^{n}X_{i,t}e_{i,t+1}\right)\Phi\right)\right)$$
$$=\frac{1}{n^2(T-k)}\sum\nolimits_{i,j=1}^{n}\mathrm{tr}\left(\Phi\Sigma_{i,j}\right)\cdot\mathrm{tr}\left(\Gamma^{-1}\Gamma_{i,j}\right)+o\left(kn^{-1}(T-k)^{-1}\right),$$

and thus (B.11) and (B.12) also hold when $k_{n,T}T^{-1/2}=o(1)$. The expectation of $H_{(2,2)}$ is solved for directly using the expression for an orthonormal basis of the polynomial time trend given in (B.1) and (B.2). Specifically,

$$\mathrm{E}\left(H_{(2,2)}\right)=\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}e_{i,T+1}e'_{i,T+1}+\frac{1}{n}\sum_{i=1}^{n}\mathbf{e}'_i\boldsymbol{\eta}\eta_{T-k+1}\eta'_{T-k+1}\boldsymbol{\eta}'\mathbf{e}_i\right)=\Sigma\left(1+\sum_{l=1}^{p}\frac{2l-1}{T-k}\frac{\prod_{r=1}^{l-1}(T-k+r)}{\prod_{r=1}^{l-1}(T-k-r)}\right).$$

For $H_{(3,3)}$, if $k_{n,T}n^{-1/2}=o(1)$ we have $\mathrm{E}\left\|\hat{\mathbf{Q}}_T-\Gamma\right\|^{2q}=O\left(k^{2q}n^{-q}\right)=o(1)$. Then using (B.10), Lemma B.3(ix), and Hölder's inequality we have

$$H_{(3,3)}=\frac{1}{n^2(T-k)^2}\sum\nolimits_{i,j=1}^{n}\mathbf{s}'_i\mathbf{M}\mathbf{X}_i\Gamma^{-1}\mathbf{X}'_j\mathbf{M}\mathbf{s}_j+o^{(d)}_{n,T}(k),$$

where $o_{n,T}^{(d)}(k)$ satisfies $\mathrm{E}\left\|o_{n,T}^{(d)}(k)\right\| = o\left(k\delta_{n,T-k}^{-2}\right)$. By application of Lemma B.3(ix) we can then establish that

$$\left|\mathrm{E}\left(\frac{1}{n^2(T-k)^2}\sum_{i,j=1}^{n}\mathbf{s}_i'\mathbf{M}\mathbf{X}_i\Gamma^{-1}\mathbf{X}_j'\mathbf{M}\mathbf{s}_j\right)\right| \leq Ck\delta_{n,T-k}^{-2}\left\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^2,$$

and hence $\max_{1\leq k\leq k_{n,T}}\left|l^{-1}(k)\,\mathrm{tr}\left(\Phi\mathrm{E}\left(H_{(3,3)}\right)\right)\right| = o(1)$ under the bound given in (B.8). Alternatively, under the $k_{n,T}T^{-1/2} = o(1)$ condition, using (B.10), the decomposition given in (B.14), and Lemma B.3(viii), we have

$$\mathrm{E}\left\|\frac{1}{n^2(T-k)^2}\sum_{i,j=1}^{n}\mathbf{s}_i'\mathbf{M}\mathbf{X}_i\Gamma^{-1}\hat{\mathbf{Q}}_T\Gamma^{-1}\mathbf{X}_j'\mathbf{M}\mathbf{s}_j - \frac{1}{n^2(T-k)^2}\sum_{t,s=k}^{T-N_0-1}\sum_{i,j,h=1}^{n}s_{i,t}X_{i,t}'\Gamma^{-1}X_{h,T}^*X_{h,T}^{*\prime}\Gamma^{-1}X_{j,s}s_{j,s}'\right\|$$
$$\leq C\frac{k}{\sqrt{n}(T-k)}\left\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^2.$$

Thus $\max_{1\leq k\leq k_{n,T}}\left|l^{-1}(k)\,\mathrm{tr}\left(\Phi\mathrm{E}\left(H_{(3,3)}\right)\right)\right| = o(1)$ under the bound given in (B.7). For $H_{(4,4)}$, first note that $\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}s_{i,T}s_{i,T}'\right) = \Lambda(k)$ and $\mathrm{tr}\left(\Phi\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{s}_i'\mathbf{P}\mathbf{s}_i\right)\right) \leq C\left\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^2$ under Lemma B.3(vii). Then using the Cauchy–Schwarz and triangle inequalities,

$$\left|\mathrm{tr}\left(\Phi\mathrm{E}\left(H_{(4,4)}\right)\right) - \left\|\boldsymbol{\alpha}-\boldsymbol{\alpha}(k)\right\|_{\Gamma\otimes\Phi}^2\right| \leq \frac{2C}{\sqrt{T-k}}\left\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^2 + \frac{C}{T-k}\left\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^2. \tag{B.15}$$

Applying the bound given in (B.7) we have

$$\max_{1\leq k\leq k_{n,T}}\left|l^{-1}(k)\left(\mathrm{tr}(\Phi\mathrm{E}\left(H_{(4,4)}\right)) - \left\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^2\right)\right| = o(1).$$

The remainder of the proof follows the same strategy as that given in the proof to Theorem 3.1 of Greenaway-McGrevy (2015) to establish that $\max_{1\leq k\leq k_{n,T}}\left|l^{-1}(k)\left(\mathrm{tr}\left(\Phi\mathrm{E}\left(H_{(r,r')}\right)\right)\right)\right| = o(1)$ for all $r \neq r'$.

B.2.2. Asymptotic Unbiasedness of QFR Estimator.

THEOREM B.4. *Let $y_{i,t}$ be generated according to (1). Under Assumptions 1, 2, and 3,*

$$\lim_{n,T\to\infty}\max_{1\leq k\leq k_{n,T}}\left|\frac{\mathrm{tr}\left(\Phi\left[\mathrm{E}\left(\hat{L}_{n,T}(k)\right)-\left(1-\frac{p^2}{T}\right)\Sigma\right]\right)}{\mathrm{tr}\left(\Phi L_{n,T}(k)\right)}-1\right| = 0,$$

for all positive semidefinite $m\times m$ matrices Φ satisfying $\|\Phi\| = 1$.

To prove the Theorem we require Lemmas B.5, B.6 and B.7 below. In the interests of brevity, for each of these Lemmas we only provide a brief sketch of the proof, which are generally straightforward applications of Theorems B.1 and B.2, elements of Lemma B.3, and the Hölder's and triangle inequalities.

LEMMA B.5. *For all k satisfying* $1 \leq k \leq k_{n,T}$,

$$\left| \mathrm{E}\left(\mathrm{tr}\left[\Phi \left(\hat{R}(k) - \Sigma \left(1 - \frac{p}{T-k_{n,T}}\right) + \frac{k}{n(T-k_{n,T})} \Pi(k) + \frac{p^2}{(T-k_{n,T})^2} \zeta(k) - \Lambda(k) \right) \right] \right) \right|$$
$$\leq C \left[\frac{k}{\delta_{n,T}^2} \left(\frac{1}{(T-k_{n,T})^{1/2}} + \|\boldsymbol{\alpha}(k) - \boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^2 \right) + \delta_{n,T}^{-1} \left(\|\boldsymbol{\alpha}(k) - \boldsymbol{\alpha}\|_{\Gamma\otimes\Phi} + \|\boldsymbol{\alpha}(k) - \boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^2 \right) \right].$$

Proof. The result follows by using the decomposition for $\hat{u}_{i,t+1}(k, k_{n,T})$ given in (B.3) to expand $\hat{R}(k) = \frac{1}{n(T-k_{n,T})} \sum_{i=1}^{n} \sum_{t=k_{n,T}}^{T-1} \hat{u}_{i,t+1}(k, k_{n,T})\, \hat{u}'_{i,t+1}(k, k_{n,T})$ into six unique terms. We then solve for the expectation of each term or bound each term using Hölder's inequality in conjunction with various elements of Lemma B.3. The leading terms to consider from this decomposition are

$$\left| \mathrm{E} \left\| \hat{\mathbf{Q}}_{(k,k_{n,T})}^{-\frac{1}{2}} \frac{1}{n(T-k_{n,T})} \sum_{i=1}^{n} \mathbf{X}_i^{(k,k_{n,T})\prime} \mathbf{M}_{T-k_{n,T}} \mathbf{e}_i^{(k_{n,T})} \right\|_{\Phi}^2 - \frac{k\mathrm{tr}(\Phi\Pi(k))}{n(T-k_{n,T})} - \frac{kp^2 \mathrm{tr}(\Phi\zeta(k))}{(T-k_{n,T})^2} \right| \leq C \left(\frac{k}{\delta_{n,T}^2} \frac{1}{(T-k_{n,T})^{1/2}} \right),$$

where the bound follows by application of Hölder's inequality, and Lemmas B.3(i), (ii), (iii), (xi), and (xii), and

$$\left| \frac{1}{n(T-k_{n,T})} \sum_{i=1}^{n} \mathrm{E} \left\| \mathbf{M}_{T-k_{n,T}} \mathbf{u}_i^{(k_{n,T},k)} \right\|_{\Phi}^2 - \mathrm{tr}\left(\Phi \left[\Sigma \left(1 - \frac{p}{T-k_{n,T}}\right) + \Lambda(k) \right] \right) \right|$$
$$\leq C \left(\frac{\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}}{\delta_{n,T}} + \frac{\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^2}{T-k_{n,T}} \right),$$

where the bound follows from Lemmas B.3(vii) and (x). ■

LEMMA B.6. *For all $q = 1, 2, \ldots$ and k satisfying* $1 \leq k \leq k_{n,T}$,

$$\mathrm{E}\left(\left| \mathrm{tr}\left[\Phi \left(\hat{\Pi}(k) - \Pi(k) \right) \right] \right|^q \right) \leq C_q \left(\frac{1}{(T-k_{n,T})^{q/2}} + \|\boldsymbol{\alpha}(k) - \boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{2q} \right). \tag{B.16}$$

Proof. The Lemma follows by Hölder's inequality and Lemmas B.3(xi) and (xii) once we establish

$$\mathrm{E} \left\| \frac{1}{n(T-k_{n,T})k} \sum_{t=k_{n,T}}^{T-1} \sum_{i,j=1}^{n} \left(\hat{u}_{i,t+1}(k, k_{n,T})\, \hat{u}'_{j,t+1}(k, k_{n,T}) \otimes \ddot{X}_{i,t}(k, k_{n,T})\, \ddot{X}'_{j,t}(k, k_{n,T}) \right) - \frac{1}{n} \sum_{i,j=1}^{n} \left(\Sigma_{i,j} \otimes \Gamma_{i,j} \right) \right\|^q$$
$$\leq C_q \left(\frac{1}{(T-k_{n,T})^{q/2}} + \|\boldsymbol{\alpha}(k) - \boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{2q} \right).$$

The bound stated above follows by using the decomposition for $\hat{u}_{i,t+1}(k, k_{n,T})$ given in (B.3) to expand out $\hat{u}_{i,t+1}(k, k_{n,T})\, \hat{u}'_{j,t+1}(k, k_{n,T})$ into sixteen separate terms, and then apply Hölder's inequality in conjunction with Theorems B.1 and B.2, various elements of Lemma B.3, and Hölder's inequality to bound each term. The leading term in this expansion is

$$\frac{1}{kn(T-k_{n,T})}\sum_{t=k_{n,T}}^{T-1}\sum_{i,j=1}^{n}\left(e_{i,t+1}e_{j,t+1}'\otimes X_{i,t}(k)X_{j,t}'(k)\right)-\frac{1}{n}\sum_{i,j=1}^{n}\left(\Sigma_{i,j}\otimes\Gamma_{i,j}\right)$$
$$=\frac{1}{kn(T-k_{n,T})}\sum_{t=k_{n,T}}^{T-1}\sum_{i,j=1}^{n}\left(\Sigma_{i,j}\otimes\left(X_{i,t}(k)X_{j,t}'(k)-\Gamma_{i,j}\right)\right)+$$
$$\frac{1}{kn(T-k_{n,T})}\sum_{t=k_{n,T}}^{T-1}\sum_{i,j=1}^{n}\left(\left(e_{i,t+1}e_{j,t+1}'-\Sigma_{i,j}\right)\otimes\Gamma_{i,j}\right)+$$
$$\frac{1}{kn(T-k_{n,T})}\sum_{t=k_{n,T}}^{T-1}\sum_{i,j=1}^{n}\left(\left(e_{i,t+1}e_{j,t+1}'-\Sigma_{i,j}\right)\otimes\left(X_{i,t}(k)X_{j,t}'(k)-\Gamma_{i,j}\right)\right).$$

By application of Theorem B.1 to the first and second terms, and Theorem B.2 to the third term, we have

$$\mathrm{E}\left\|\left(\frac{1}{kn(T-k_{n,T})}\sum_{t=k_{n,T}}^{T-1}\sum_{i,j=1}^{n}\left(e_{i,t+1}e_{j,t+1}'\otimes X_{i,t}(k)X_{j,t}'(k)\right)\right)-\frac{1}{nk}\sum_{i,j=1}^{n}\left(\Sigma_{i,j}\otimes\Gamma_{i,j}\right)\right\|^{q}\leq\frac{C_q}{(T-k_{n,T})^{\frac{q}{2}}}.$$

The result follows because the remaining fifteen terms in the expansion satisfy the bound given on the right hand side of (B.16). ■

LEMMA B.7. *For all $q=1,2,\ldots$ and k satisfying $1\leq k\leq k_{n,T}$,*

$$\mathrm{E}\left(\left|\mathrm{tr}\left(\Phi\left(\hat{\zeta}(k)-\zeta(k)\right)\right)\right|^{q}\right)\leq C_q\left(\frac{1}{\delta_{n,T}^{q}}+\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{2q}+\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{q}\right).$$

Proof. The stated result follows from Hölder's inequality and Lemmas B.3(xi) and (xii) once we establish that

$$\mathrm{E}\left\|\hat{\boldsymbol{\xi}}_{(k,l)}-\boldsymbol{\xi}(k)\right\|^{2q}\leq C_q\left(\frac{k^q}{n^q(T-l)^q}+\frac{k^q}{(T-l)^{2q}}+k^q\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Phi\otimes\Gamma}^{2q}\right)\tag{B.17}$$

for $l=k,k+1,\ldots,k_{n,T}$. The above bound follows by application of the triangle and Hölder's inequalities and various elements of Lemma B.3, given that

$$\left\|\sum_{s=1}^{\infty}\alpha_s-\sum_{s=1}^{k}\alpha_s(k)\right\|^{2q}\leq C_q\left(\mathrm{E}\left\|\sum_{s=k+1}^{\infty}\alpha_s'x_{i,t-s+1}\right\|^{2}\right)^{q}=C_q\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{2q},$$

(also see (3.5) in Cheng, Ing, and Yu, 2014) and

$$\mathrm{E}\left\|(I_{mk}-\hat{\mathbf{A}}_{(k,l)})^{-1}\right\|^{q}=O(1).\tag{B.18}$$

To demonstrate that (B.18) holds, we first define the $(k+1)m\times(k+1)m$ matrix

$$\hat{\mathbf{Q}}_{1|(k+1,l)}:=\frac{1}{n(T-l)}\sum_{i=1}^{n}\mathbf{X}_{i|1}^{(k+1,l)\prime}\mathbf{M}_{T-l+1}\mathbf{X}_{i|1}^{(k+1,l)},$$

where $\mathbf{X}_{i|1}^{(k+1,l)} := \left[X_{i,k_{nT}}(k+1) : \cdots : X_{i,T}(k+1)\right]'$, such that

$$I_{mk} - \hat{\mathbf{A}}_{(k,l)}' = \hat{\mathbf{Q}}_{(k,l)}^{-1}\mathbf{J}_{m(k+1),mk}'\hat{\mathbf{Q}}_{(k+1,l)}\mathbf{D}_{k+1} + o_{n,T}^{(e)}(k),$$

for an appropriately defined $(k+1)m \times km$ first-differencing matrix $\mathbf{D}_{k+1}$, and where $o_{n,T}^{(e)}(k)$ is a negligible term satisfying $\mathrm{E}\left\|o_{n,T}^{(e)}(k)\right\|^q = o(1)$. Then (B.18) follows from the fact that the minimum eigenvalue of $\hat{\mathbf{Q}}_{1|(k+1,l)}$ is bounded below by zero in expectation under Lemma B.3(xii), and because the km-th largest singular value of the $(k+1)m \times m(k+1)$ matrix $\mathbf{D}_{k+1}\mathbf{J}_{m(k+1),mk}'$ is strictly positive. ■

To conserve space let $\hat{l}(k) := \mathrm{tr}\left(\Phi\hat{L}_{n,T}(k)\right)$. Note that

$$\frac{1}{n(T-k_{n,T})}\sum_{i=1}^{n}\sum_{t=k_{n,T}}^{T-1}\left\|\hat{u}_{i,t+1}\left(k,k_{n,T}\right)\right\|_{\Phi}^2 = \frac{1}{n(T-k_{n,T})}\sum_{i=1}^{n}\left\|\mathbf{M}_{T-k_{n,T}}\mathbf{u}_i^{(k,k_{n,T})}\right\|_{\Phi}^2 - \left\|\hat{\mathbf{Q}}_{(k,k_{n,T})}^{-\frac{1}{2}}\frac{1}{n(T-k_{n,T})}\sum_{i=1}^{n}\mathbf{X}_i^{(k,k_{n,T})\prime}\mathbf{M}_{T-k_{n,T}}\mathbf{u}_i^{(k,k_{n,T})}\right\|_{\Phi}^2$$

so that we can decompose

$$\hat{l}(k) - l(k) - \left(1+\frac{p^2}{T}\right)\mathrm{tr}(\Phi\Sigma) \tag{B.19}$$
$$= \underbrace{\frac{k\mathrm{tr}(\Phi\Pi(k))}{n(T-k_{n,T})} + \frac{p^2 k\mathrm{tr}(\Phi\zeta(k))}{(T-k_{n,T})^2} - \left\|\hat{\mathbf{Q}}_{(k,k_{n,T})}^{-\frac{1}{2}}\frac{1}{n(T-k_{n,T})}\sum_{i=1}^{n}\mathbf{X}_i^{(k,k_{n,T})\prime}\mathbf{M}_{T-k_{n,T}}\mathbf{u}_i^{(k,k_{n,T})}\right\|_{\Phi}^2}_{=:\imath_1(k)} +$$
$$\underbrace{\left(\frac{kp^2}{(T-k_{n,T})^2} + \frac{kp^2}{(T-k)^2}\right)\mathrm{tr}\left[\Phi\left(\hat{\zeta}(k) - \zeta(k)\right)\right] + \frac{k}{n}\left(\frac{1}{T-k_{n,T}} + \frac{1}{T-k}\right)\mathrm{tr}\left[\Phi\left(\hat{\Pi}(k) - \Pi(k)\right)\right]}_{=:\imath_2(k)} +$$
$$\underbrace{\frac{1}{n(T-k_{n,T})}\sum_{i=1}^{n}\left\|\mathbf{M}_{T-k_{n,T}}\mathbf{u}_i^{(k,k_{n,T})}\right\|_{\Phi}^2 - \mathrm{tr}(\Phi\Lambda(k)) - \left(1-\frac{p}{T-k_{n,T}}\right)\mathrm{tr}(\Phi\Sigma)}_{=:\imath_3(k)} +$$
$$\underbrace{\left(\frac{p}{T-k_{n,T}} + \sum_{l=1}^{p}\frac{2l-1}{T-k}\frac{\prod_{r=0}^{l-1}(T-k-r)}{\prod_{r=0}^{l-1}(T-k+r)}\right)\mathrm{tr}\left[\Phi\left(\hat{R}(k)\frac{T-k_{n,T}}{T-k_{n,T}-p} - \Sigma\right)\right]}_{=:\imath_4(k)}.$$

By Hölder's inequality and Lemmas B.3(i), (ii), (ix), (xi), and (xii), we can establish that

$$|\mathrm{E}(\imath_1(k))| = \left|\mathrm{E}\left\|\hat{\mathbf{Q}}_{(k,k_{n,T})}^{-\frac{1}{2}}\frac{1}{n(T-k_{n,T})}\sum_{i=1}^{n}\mathbf{X}_i^{(k,k_{n,T})\prime}\mathbf{M}_{T-k_{n,T}}\mathbf{u}_i^{(k,k_{n,T})}\right\|_{\Phi}^2 - \frac{k\mathrm{tr}(\Phi\Pi(k))}{n(T-k_{n,T})} - \frac{kp^2\mathrm{tr}(\Phi\zeta(k))}{(T-k_{n,T})^2}\right|$$
$$\leq C\left(\frac{k}{\delta_{n,T}^2}\frac{1}{(T-k_{n,T})^{1/2}} + \frac{k\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^2}{\delta_{n,T}^2} + \frac{k}{\delta_{n,T}^2}\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}\right).$$

Next, by Lemmas B.6 and B.7 and the triangle inequality we have

$$|\mathrm{E}(\imath_2(k))| \leq C\frac{k}{\delta_{n,T}^2}\left(\frac{1}{(T-k_{n,T})^{1/2}}+\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^2+\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}\right).$$

Next, by Hölder's inequality and Lemmas B.3(iv), (v), (vi), (vii), and (x), we have

$$|\mathrm{E}(\imath_3(k))| \leq C\left(\delta_{n,T}^{-1}\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^2+\delta_{n,T}^{-1}\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}\right).$$

Finally, by Lemma B.5,

$$|\mathrm{E}(\imath_4(k))| \leq C\frac{1}{T-k_{n,T}}\left(\frac{k}{\delta_{n,T}^2}+\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^2+\frac{1}{T-k_{n,T}}\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}\right).$$

Using the four inequalities above, together with the bounds and limits given in (B.6), (B.7), (B.8), and (B.9), we have

$$\max_{1\leq k\leq k_{n,T}}\left|l^{-1}(k)\left(\hat{l}(k)-l(k)\right)\right| = o(1),$$

completing the proof. Note that the above limit holds even if we were to relax $k_{n,T}$ to be a sequence satisfying $k_{n,T}n^{-1/2}T^{-1/2}=o(1)$ and $k_{n,T}T^{-1}=o(1)$.

B.2.3. Proof of Theorem 3.2 (*Asymptotic Efficiency*). Throughout this subsection we use $k^*=k^*_{n,T}$ for brevity. We first establish Lemmas B.8 and B.9 below.

LEMMA B.8. (*i*) *For all integers $q>1$ there exists a finite constant C_q such that*

$$\mathrm{E}\left(\left|\hat{l}(k)-l(k)-\hat{l}(k^*)+l(k^*)\right|^q\right)$$
$$\leq C_q\left(\frac{k^q+k^{*q}}{\delta_{n,T}^{2q}}\frac{1}{(T-k_{n,T})^{\frac{q}{2}}}+\frac{k^q+k^{*q}}{n^{\frac{q}{2}}(T-k_{n,T})^{\frac{3q}{2}}}+\frac{\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{2q}+\|\boldsymbol{\alpha}(k^*)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{2q}}{\delta_{n,T}^q}+\frac{\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{q}+\|\boldsymbol{\alpha}(k^*)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{q}}{\delta_{n,T}^q}\right)$$

(*ii*) *For all $r>1$ and q such that $q>\upsilon^{-1}2(r+1)$,*

$$\sum_{k=1}^{k_{n,T}}\delta_{n,T}^{2r}l^r(k)\mathrm{E}\left(\left|\frac{\hat{l}(k)-l(k)-\hat{l}(k^*)+l(k^*)}{l(k)}\right|^q\right)=o(1).$$

Proof. For brevity, within this proof we let $\hat{\mathbf{Q}}:=\hat{\mathbf{Q}}_{(k,k_{n,T})}$, $\mathbf{X}_i:=\mathbf{X}_i^{(k,k_{n,T})}$, $\mathbf{u}_i^{(k)}:=\mathbf{u}_i^{(k,k_{n,T})}$, $\mathbf{s}_i^{(k)}:=\mathbf{s}_i^{(k,k_{n,T})}$, $\mathbf{e}_i:=\mathbf{e}_i^{(k_{n,T})}$, $\mathbf{M}:=\mathbf{M}_{T-k_{n,T}}$, $\mathbf{P}:=\mathbf{P}_{T-k_{n,T}}$, $\Gamma:=\Gamma(k)$ and $\Gamma_{i,j}:=\Gamma_{i,j}(k)$. Applying the triangle inequality to the decomposition in (B.19), we have

$$\mathrm{E}\left(\left|\hat{l}(k)-l(k)-\left(\hat{l}(k^*)-l(k^*)\right)\right|^q\right)\leq \tag{B.20}$$
$$C_q\left(\mathrm{E}\left|\imath_1(k^*)\right|^q+\mathrm{E}\left|\imath_1(k)\right|^q+\mathrm{E}\left|\imath_2(k^*)\right|^q+\mathrm{E}\left|\imath_2(k)\right|^q+\mathrm{E}\left|\imath_3(k)-\imath_3(k^*)\right|^q+\mathrm{E}\left|\imath_4(k)-\imath_4(k^*)\right|^q\right).$$

We now bound each expectation on the right hand side of the inequality. By Theorem B.2 and Lemma B.3(i), we can establish that

$$\mathrm{E}\left|\mathrm{tr}\left(\frac{1}{n^2(T-k_{n,T})^2}\sum_{i,j=1}^n\mathbf{e}_i'\mathbf{X}_i\Gamma^{-1}\mathbf{X}_j'\mathbf{e}_j\Phi\right)-\frac{1}{n^2(T-k_{n,T})^2}\sum_{i,j=1}^n\mathrm{tr}(\Phi\Sigma_{i,j})\,\mathrm{tr}\left(\Gamma^{-1}\Gamma_{i,j}\right)\right|^q\leq C_q\frac{k^q}{n^q(T-k_{n,T})^{\frac{3q}{2}}}.$$

This result, together with Hölder's inequality, Lemmas B.3(i), (ii), (ix), (xi), and (xii), we can establish that

$$\begin{aligned}\mathrm{E}\left|\iota_1(k)\right|^q &= \mathrm{E}\left|\left\|\hat{\mathbf{Q}}^{-1/2}\frac{1}{n(T-k_{n,T})}\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{M}\mathbf{u}_i^{(k)}\right\|_{\Phi}^2-\frac{k}{n(T-k_{n,T})}\mathrm{tr}(\Phi\Pi(k))-\frac{kp^2}{(T-k_{n,T})^2}\mathrm{tr}(\Phi\zeta(k))\right|^q\\ &\le C_q\left(\frac{k^q}{\delta_{n,T}^{2q}}\frac{1}{(T-k_{n,T})^{q/2}}+\frac{k^q}{n^{q/2}(T-k_{n,T})^{3q/2}}+\frac{k^q\left\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{2q}}{\delta_{n,T}^{2q}}+\frac{k^q\left\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{q}}{\delta_{n,T}^{2q}}\right). \end{aligned}\qquad \textbf{(B.21)}$$

Next, by Lemmas B.6 and B.7 and Hölder's inequality, we have

$$\mathrm{E}\left|\iota_2(k)\right|^q \le C_q\frac{k^q}{\delta_{n,T}^{2q}}\left(\frac{1}{(T-k_{n,T})^{q/2}}+\left\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{2q}+\left\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{q}\right). \qquad \textbf{(B.22)}$$

Turning to $\iota_3(k)-\iota_3\left(k^*\right)$, we first note that

$$\frac{1}{n(T-k_{n,T})}\sum_{i=1}^{n}\left\|\mathbf{M}\mathbf{u}_i^{(k)}\right\|_{\Phi}^2=\frac{1}{n(T-k_{n,T})}\sum_{i=1}^{n}\left(\left\|\mathbf{M}\mathbf{s}_i^{(k)}\right\|_{\Phi}^2+\left\|\mathbf{M}\mathbf{e}_i\right\|_{\Phi}^2+\mathrm{tr}\left(\Phi\mathbf{e}_i'\mathbf{M}\mathbf{s}_i^{(k)}\right)+\mathrm{tr}\left(\Phi\mathbf{s}_i^{(k)\prime}\mathbf{M}\mathbf{e}_i\right)\right).$$

Then by Lemmas B.3(vi), (vii), and (x), we have

$$\begin{aligned}\mathrm{E}\left|\iota_3(k)-\iota_3\left(k^*\right)\right|^q &\le C_q\frac{1}{n(T-k_{n,T})}\sum_{i=1}^{n}\left(\mathrm{E}\left|\left\|\mathbf{M}\mathbf{s}_i^{(k)}\right\|_{\Phi}^2-\mathrm{tr}(\Phi\Lambda(k))\right|^q+2\mathrm{E}\left(\left|\mathrm{tr}\left(\Phi\mathbf{e}_i'\mathbf{M}\mathbf{s}_i^{(k)}\right)\right|^q\right)+\right.\\ &\qquad\left.\mathrm{E}\left|\left\|\mathbf{M}\mathbf{s}_i^{(k^*)}\right\|_{\Phi}^2-\mathrm{tr}\left(\Phi\Lambda\left(k^*\right)\right)\right|^q+2\mathrm{E}\left(\left|\mathrm{tr}\left(\Phi\mathbf{e}_i'\mathbf{M}\mathbf{s}_i^{(k^*)}\right)\right|^q\right)\right)\\ &\le C_q\left(\frac{\left\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{2q}+\left\|\boldsymbol{\alpha}\left(k^*\right)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{2q}}{\delta_{n,T}^{2q}}+\frac{\left\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{q}+\left\|\boldsymbol{\alpha}\left(k^*\right)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{q}}{\delta_{n,T}^{q}}\right). \end{aligned}\qquad \textbf{(B.23)}$$

Finally for $\mathrm{E}\left|\iota_4(k)-\iota_4\left(k^*\right)\right|^q$ we first note that

$$\begin{aligned}\mathrm{tr}\left[\Phi\left(\hat{R}(k)-\hat{R}\left(k^*\right)\right)\right] &= \frac{1}{n(T-k_{n,T}-p)}\sum_{i=1}^{n}\sum_{t=k_{n,T}}^{T-1}\left(\left\|\hat{u}_{i,t+1}\left(k,k_{n,T}\right)\right\|_{\Phi}^2-\left\|\ddot{e}_{i,t+1}\left(k_{n,T}\right)\right\|_{\Phi}^2\right)-\\ &\quad\frac{1}{n(T-k_{n,T}-p)}\sum_{i=1}^{n}\sum_{t=k_{n,T}}^{T-1}\left(\left\|\hat{u}_{i,t+1}\left(k^*,k_{n,T}\right)\right\|_{\Phi}^2-\left\|\ddot{e}_{i,t+1}\left(k_{n,T}\right)\right\|_{\Phi}^2\right).\end{aligned}$$

Then by Hölder's inequality and Lemmas B.3(i), (ii), (vi), (vii), (ix), (xi) and (xii), we can establish that

$$\mathrm{E}\left(\left|\mathrm{tr}\left[\Phi\left(\hat{R}(k)-\hat{R}\left(k^*\right)\right)\right]\right|^q\right)\le C_q\left(\begin{array}{c}\frac{k^q+k^{*q}}{\delta_{n,T}^{2q}}+\left\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{2q}+\left\|\boldsymbol{\alpha}\left(k^*\right)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{2q}+\\ \frac{\left\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{q}}{\delta_{n,T}^{q}}+\frac{\left\|\boldsymbol{\alpha}\left(k^*\right)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{q}}{\delta_{n,T}^{q}}\end{array}\right),$$

and thus we have

$$\mathrm{E}\left|\iota_4(k)-\iota_4\left(k^*\right)\right|^q\le C_q\left(\frac{k^q+k^{*q}}{\delta_{n,T}^{3q}}+\frac{\left\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{2q}}{\delta_{n,T}^{q}}+\frac{\left\|\boldsymbol{\alpha}\left(k^*\right)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{2q}}{\delta_{n,T}^{q}}+\frac{\left\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{q}}{\delta_{n,T}^{2q}}+\frac{\left\|\boldsymbol{\alpha}\left(k^*\right)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{q}}{\delta_{n,T}^{2q}}\right). \qquad \textbf{(B.24)}$$

Part (i) of the Lemma then follows by substituting the bounds given in (B.21), (B.22), (B.23) and (B.24) into (B.20). Turning to part (ii) of the Lemma, since $\delta_{n,T-k}^{-2(q-r)}l^{-(q-r)}(k)\le C_{q-r}k^{-(q-r)}$, $\left\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{2q}\le l^q(k)$, $k^q\delta_{n,T}^{-2q}\left\|\Sigma^{1/2}\right\|_{\Phi}^{q}\left\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{q}\le l^q(k)$, $l^r(k)\le C_r\left(k^r\delta_{n,T-k}^{-2r}+k^{-r}\right)$ and $l\left(k^*\right)\le l(k)$, we have

$$
\begin{aligned}
&\sum_{k=1}^{k_{n,T}} \delta_{n,T}^{2r} \tfrac{l^r(k)}{l^q(k)} \left(\mathrm{E}\left|\iota_1(k^*)\right|^q + \mathrm{E}\left|\iota_1(k)\right|^q + \mathrm{E}\left|\iota_2(k)\right|^q + \mathrm{E}\left|\iota_2(k^*)\right|^q + \mathrm{E}\left|\iota_3(k) - \iota_3(k^*)\right|^q + \mathrm{E}\left|\iota_4(k) - \iota_4(k^*)\right|^q \right) \\
&\le C_q \sum_{k=1}^{k_{n,T}} \left(\frac{k^q + k^{*q}}{\delta_{n,T}^{2q-2r} l^{q-r}(k) \left(T - k_{n,T}\right)^{\frac{q}{2}}} + \frac{\delta_{n,T}^{2q}\left(k^q + k^{*q}\right)}{\delta_{n,T}^{2q-2r} l^{q-r}(k) n^{\frac{q}{2}} \left(T - k_{n,T}\right)^{\frac{3q}{2}}} + \right. \\
&\qquad \left. \frac{l^r(k)}{\delta_{n,T}^{q-2r}} \frac{\left\|\boldsymbol{\alpha}(k) - \boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{2q} + \left\|\boldsymbol{\alpha}(k^*) - \boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{2q}}{l^q(k)} + \frac{\left\|\boldsymbol{\alpha}(k) - \boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{q} + \left\|\boldsymbol{\alpha}(k^*) - \boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{q}}{\delta_{n,T}^{q-2r} l^{q-r}(k)} \right) \\
&\le C_q \sum_{k=1}^{k_{n,T}} \left(C_{q-r} \left[\frac{k^r + k^{*r}}{\left(T - k_{n,T}\right)^{\frac{q}{2}}} + \frac{\delta_{n,T}^{2q}\left(k^r + k^{*r}\right)}{n^{\frac{q}{2}} \left(T - k_{n,T}\right)^{\frac{3q}{2}}} \right] + C_r \left[\frac{k^r}{\delta_{n,T}^q} + \frac{1}{k^r \delta_{n,T}^{q-2r}} + \frac{k^r}{\delta_{n,T}^q} + \frac{1}{k^{*r} \delta_{n,T}^{q-2r}} \right] \right) + \\
&\quad C_q \sum_{k=1}^{k_{n,T}} C_{q-2r} k^{-\frac{q}{2}+r} + C_q \left[\sum_{k=1}^{k^*} k^{*-\left(\frac{q}{2}-r\right)} + k^{*-\left(\frac{q}{2}-r\right)\underline{\pi}} \sum_{k=k^*+1}^{k_{n,T}} k^{-\left(\frac{q}{2}-r\right)\left(1-\underline{\pi}\right)} + \sum_{k=1}^{k_{n,T}} k^{*-\frac{1}{2}(q-2r)} \right] = o(1),
\end{aligned}
$$

for some $\underline{\pi}$ satisfying $0 \le \underline{\pi} < \left(\frac{q}{2} - r\right)^{-1}$. Each sum is $o(1)$ for all $q > \upsilon^{-1} 2(r+1)$ and $r > 1$ because $k^* \to \infty$, $\sum_{k=1}^{k_{n,T}} k^r = O\left(k_{n,T}^{r+1}\right)$, $\sum_{k=1}^{k_{n,T}} k^{-r} = O(1)$, $\sum_{k=1}^{k_{n,T}} k^{-(q-r)} = O(1)$, $\sum_{k=1}^{k_{n,T}} k^{-(q/2-r)} = O(1)$ and $k_{n,T}^{2r} \delta_{n,T}^{-q} = o(1)$. Part (ii) of the Lemma thus follows. ■

LEMMA B.9. *For all integers $q > 1$ and k_1, k_2 satisfying $1 \le k_1, k_2 \le k_{n,T}$,*

$$
\mathrm{E}\left(\frac{1}{n} \sum_{i=1}^{n} \left\| \left(\hat{u}_{i,T+1}(k_2, k_2) - \hat{u}_{i,T+1}(k_1, k_1) \right) \right\|_{\Phi}^2 \right)^q \le C_q \left(\left| \frac{k_2 - k_1}{\delta_{n,T}^2} \right|^q + \left| l(k_2) - l(k_1) \right|^q \right). \tag{B.25}
$$

Proof. For brevity, within this proof we let $\mathbf{X}_i^{(k)} := \mathbf{X}_i^{(k,k)}$, $\mathbf{u}_i^{(k)} := \mathbf{u}_i^{(k,k)}$, $\mathbf{s}_i^{(k)} := \mathbf{s}_i^{(k,k)}$ and $\hat{\mathbf{Q}}_{(k)} := \hat{\mathbf{Q}}_{(k,k)}$. Similarly, let $\bar{e}_{i,T}(k) := \mathbf{e}_i^{(k)\prime} \boldsymbol{\eta}_{T-k} \eta_{T-k+1}^{(T-k)}$, $\bar{s}_{i,T}(k) := \mathbf{s}_i^{(k)\prime} \boldsymbol{\eta}_{T-k} \eta_{T-k+1}^{(T-k)}$, and $\ddot{X}_{i,T}(k) := X_{i,T}(k) - \bar{X}_{i,T}(k)$ for $\bar{X}_{i,T}(k) := \mathbf{X}_i^{(k)\prime} \boldsymbol{\eta}_{T-k} \eta_{T-k+1}^{(T-k)}$. Then, for each k we have

$$
\hat{u}_{i,T+1}(k,k) = \ddot{X}_{i,T}'(k) \hat{\mathbf{Q}}_{(k)}^{-1} \tfrac{1}{n(T-k)} \textstyle\sum_{j=1}^{n} \mathbf{X}_j^{(k)\prime} \mathbf{M}_{T-k} \mathbf{u}_j^{(k)} - e_{i,T+1} + \bar{e}_{i,T}(k) - s_{i,T}(k) + \bar{s}_{i,T}(k).
$$

We can then decompose

$$
\begin{aligned}
\hat{u}_{i,T+1}(k_2,k_2) - \hat{u}_{i,T+1}(k_1,k_1) = &\underbrace{\left(\ddot{X}_{i,T}(k_2) - \underline{\ddot{X}}_{i,T}(k_1) \right)' \hat{\mathbf{Q}}_{(k_2)}^{-1} \tfrac{1}{n(T-k_2)} \sum_{j=1}^{n} \underline{\mathbf{X}}_j^{(k_2)\prime} \underline{\mathbf{M}}_{T-k_2} \underline{\mathbf{u}}_j^{(k_2)}}_{=:\chi_{1,i}(k_2,k_1)} + \\
&\underbrace{\underline{\ddot{X}}_{i,T}'(k_1) \underline{\hat{\mathbf{Q}}}_{(k_1)}^{-1} \tfrac{1}{n} \sum_{j=1}^{n} \left(\tfrac{1}{T-k_2} \underline{\mathbf{X}}_j^{(k_2)\prime} \underline{\mathbf{M}}_{T-k_2} \underline{\mathbf{u}}_j^{(k_2)} - \tfrac{1}{T-k_1} \underline{\mathbf{X}}_j^{(k_1)\prime} \mathbf{M}_{T-k_1} \mathbf{u}_j^{(k_1)} \right)}_{=:\chi_{2,i}(k_2,k_1)} + \\
&\underbrace{\underline{\ddot{X}}_{i,T}'(k_1) \left(\hat{\mathbf{Q}}_{(k_2)}^{-1} - \underline{\hat{\mathbf{Q}}}_{(k_1)}^{-1} \right) \tfrac{1}{n(T-k_2)} \sum_{j=1}^{n} \underline{\mathbf{X}}_j^{(k_2)\prime} \underline{\mathbf{M}}_{T-k_2} \underline{\mathbf{u}}_j^{(k_2)}}_{=:\chi_{3,i}(k_2,k_1)} + \\
&\underbrace{\left[\bar{e}_{i,T}(k_2) - \bar{e}_{i,T}(k_1) \right]}_{=:\chi_{4,i}(k_2,k_1)} + \underbrace{\left[s_{i,T}(k_1) - \bar{s}_{i,T}(k_1) - s_{i,T}(k_2) + \bar{s}_{i,T}(k_2) \right]}_{=:\chi_{5,i}(k_2,k_1)},
\end{aligned}
$$

where $\underline{\hat{\mathbf{Q}}}_{(k_1)}^{-1}$ denotes a $k_2 \times k_2$ matrix with $\hat{\mathbf{Q}}_{(k_1)}^{-1}$ in the upper left $k_1 \times k_1$ block and zeros elsewhere; $\underline{\ddot{X}}_{i,T}(k_1) := \left(\ddot{X}_{i,T}'(k_1), 0, \ldots, 0\right)'$ is an $mk_2 \times 1$ vector; $\underline{\mathbf{M}}_{T-k_2}$ is a $(T-k_1) \times (T-k_1)$ matrix with $\mathbf{M}_{T-k_2}$ in the upper left $(T-k_2) \times (T-k_2)$ block and zeros elsewhere; $\underline{\mathbf{X}}_i^{(k_2)} := \left[\mathbf{X}_i^{(k_2)\prime} : \mathbf{0}_{k_2 m \times (k_2-k_1)}\right]'$ is a $(T-k_1) \times k_2 m$ matrix; $\underline{\mathbf{X}}_i^{(k_1)} := \left[\mathbf{X}_i^{(k_1)} : \mathbf{0}_{(T-k_1)\times(k_2-k_1)m}\right]$ is a $(T-k_1) \times k_2 m$ matrix; and $\underline{\mathbf{u}}_i^{(k_2)} := \left[\mathbf{u}_i^{(k_2)\prime} : \mathbf{0}_{m\times(k_2-k_1)}\right]'$ is a $(T-k_1) \times m$ matrix. The proof then proceeds by showing that $\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^n \left\|\chi_{s,i}(k_2,k_1)\right\|_\Phi^2\right)^q$ satisifies the bound given in (B.25) for each $s = 1, 2, \ldots, 5$. First we decompose

$$\chi_{1,i}(k_2,k_1) = \underline{x}_{i,T}'(k_2,k_1)\mathbf{J}_{k_2,k_2-k_1}'\hat{\mathbf{Q}}_{(k_2)}^{-1}\frac{1}{n(T-k_2)}\sum_{j=1}^n \underline{\mathbf{X}}_j^{(k_2)\prime}\underline{\mathbf{M}}_{T-k_2}\underline{\mathbf{u}}_j^{(k_2)} + \left(\bar{X}_{i,T}(k_2) - \underline{\bar{X}}_{i,T}(k_1)\right)'\hat{\mathbf{Q}}_{(k_2)}^{-1}\frac{1}{n(T-k_2)}\sum_{j=1}^n \underline{\mathbf{X}}_j^{(k_2)\prime}\underline{\mathbf{M}}_{T-k_2}\underline{\mathbf{u}}_j^{(k_2)}, \quad \textbf{(B.26)}$$

where $\underline{x}_{i,t}(k_2,k_1) := \left(x_{i,t-k_1}', \ldots, x_{i,t-k_2+1}'\right)'$ is a $m(k_2-k_1) \times 1$ vector and $\underline{\bar{X}}_{i,T}(k_1) := \left(\bar{X}_{i,T}'(k_1), 0, \ldots, 0\right)'$ is a $mk_2 \times 1$ is a vector. First we assume that (7) holds under Assumption 3. By Theorem B.1 we can establish that

$$\mathrm{E}\left(\frac{1}{n}\sum_i \left\|\underline{x}_{i,T}(k_2,k_1)\right\|^2\right)^q = \mathrm{E}\left(\rho_{\max}^q\left(\frac{1}{n}\sum_i \underline{x}_{i,T}(k_2,k_1)\underline{x}_{i,T}'(k_2,k_1)\right)\right) \le C_q\left(\frac{1}{n^{q/2}}(k_2-k_1)^q + 1\right).$$

Meanwhile, by Lemmas B.3(iii), (ix), (xi), (xii), and Theorem B.1, we have

$$\mathrm{E}\left\|\mathbf{J}_{k_2,k_2-k_1}'\hat{\mathbf{Q}}_{(k_2)}^{-1}\frac{1}{n(T-k_2)}\sum_{j=1}^n \underline{\mathbf{X}}_j^{(k_2)\prime}\underline{\mathbf{M}}_{T-k_2}\mathbf{u}_j(k_2)\right\|_\Phi^{2q} \le C_q\left(\left|\frac{k_2-k_1}{\delta_{n,T}^2}\right|^q + \left|\frac{k_2^2}{\delta_{n,T}^2}\right|^q\left|\frac{k_2-k_1}{\delta_{n,T}^2}\right|^q\left|\frac{1}{k_2-k_1}\right|^q\right), \quad \textbf{(B.27)}$$

and thus the first term in (B.26) satisfies

$$\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^n\left\|\underline{x}_{i,T}'(k_2,k_1)\mathbf{J}_{k_2,k_2-k_1}'\hat{\mathbf{Q}}_{(k_2)}^{-1}\frac{1}{n(T-k_2)}\sum_{j=1}^n \underline{\mathbf{X}}_j^{(k_2)\prime}\underline{\mathbf{M}}_{T-k_2}\mathbf{u}_j(k_2)\right\|_\Phi^2\right)^q \le C_q\left(\left|\frac{k_2-k_1}{\delta_{n,T}^2}\right|^q\right)$$

provided $kn^{-1/2} = o(1)$. If (8) holds instead of (7) under Assumption 3, using the decomposition given in (B.13) and following a similar strategy as given in the proof to Theorem 3.1, we can decompose

$$\underline{x}_{i,T}'(k_2,k_1)\mathbf{J}_{k_2,k_2-k_1}'\hat{\mathbf{Q}}_{(k_2)}^{-1}\frac{1}{n(T-k_2)}\sum_{j=1}^n \underline{\mathbf{X}}_j^{(k_2)\prime}\underline{\mathbf{M}}_{T-k_2}\underline{\mathbf{u}}_j^{(k_2)}$$
$$= \underline{x}_{i,T}^{*\prime}(k_2,k_1)\mathbf{J}_{k_2,k_2-k_1}'\Gamma^{-1}(k_2)\frac{1}{n(T-k_2)}\sum_{s,t=k_2}^{T-N_0-1}\sum_{j=1}^n X_{j,t}(k_2)e_{j,t+1}' + o_{n,T}^{(f)}(k_2),$$

where we assume that $k_2 < \sqrt{T-k_2}$ without loss of generality; N_0 is the largest integer less than $\sqrt{T-k_2}$; the term $o_{n,T}^{(f)}(k_2)$ satisfies $\mathrm{E}\left\|o_{n,T}^{(f)}(k_2)\right\|^{2q} \le C_q(k_2-k_1)^q\delta_{n,T}^{-2q}$; $\underline{x}_{i,t}^*(k_2,k_1) := \left(x_{i,t-k_1}^{*\prime}, \ldots, x_{i,t-k_2+1}^{*\prime}\right)'$ for $x_{i,t}^*$ defined in (B.13); and

$$\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^n\left\|\underline{x}_{i,T}^{*\prime}(k_2,k_1)\mathbf{J}_{k_2,k_2-k_1}'\Gamma^{-1}(k_2)\frac{1}{n(T-k_2)}\sum_{s,t=k_2}^{T-N_0-1}\sum_{j=1}^n X_{j,t}(k_2)e_{j,t+1}'\right\|_\Phi^2\right)^q \le C_q(k_2-k_1)^q\delta_{n,T}^{-2q}.$$

Turning to the second term in (B.26), we first define

$$\eta_{\boldsymbol{\Delta},t}^{(T,k_2,k_1)} := \begin{cases} \eta_t^{(T-k_1)\prime}\eta_{T-k_1+1}^{(T-k_1)} & \text{for } t = 1,\ldots,k_2-k_1 \\ \eta_{t-k_2+k_1}^{(T-k_2)\prime}\eta_{T-k_2+1}^{(T-k_2)} - \eta_t^{(T-k_1)\prime}\eta_{T-k_1+1}^{(T-k_1)} & \text{for } t = k_2-k_1+1,\ldots,T-k_1 \end{cases}$$

where we can straightforwardly establish that there exists finite C such that

$$\sum_{t=1}^{T-k_2}\left(\eta_{\boldsymbol{\Delta},t}^{(T,k_2,k_1)}\right)^2 \le C\left(k_2-k_1\right)\left(T-k_1\right)^{-2}.$$

We can then decompose

$$\bar{X}_{i,T}\left(k_2\right) - \underline{\bar{X}}_{i,T}\left(k_1\right) = \underline{\mathbf{X}}_i^{(k_1)\prime}\boldsymbol{\eta}_{\boldsymbol{\Delta}} + \underline{\mathbf{x}}_i^{(k_2,k_1)\prime}\boldsymbol{\eta}_{T-k_2}\eta_{T-k_2+1}^{(T-k_2)},$$

where $\boldsymbol{\eta}_{\boldsymbol{\Delta}} := \left(\eta_{\boldsymbol{\Delta},1}^{(T,k_2,k_1)},\ldots,\eta_{\boldsymbol{\Delta},T-k_1}^{(T,k_2,k_1)}\right)'$ and

$$\underline{\mathbf{x}}_i^{(k_2,k_1)} := \begin{bmatrix} & \underline{x}_{i,k_2}'\left(k_1,k_2\right) \\ \mathbf{0}_{(T-k_2)\times mk_1} & \vdots \\ & \underline{x}_{i,T-1}'\left(k_1,k_2\right) \\ \multicolumn{2}{c}{\mathbf{0}_{(k_2-k_1)\times m(k_2-k_1)}} \end{bmatrix}.$$

By the Cauchy–Schwarz and triangle inequalities and Theorem B.1 we then have

$$\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\left(\bar{X}_{i,T}\left(k_2\right) - \underline{\bar{X}}_{i,T}\left(k_1\right)\right)\right\|_{\Phi}^{2}\right)^q \le C_q\left(k_2-k_1\right)^q\left(T-k_1\right)^{-q}.$$

These results imply that

$$\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\chi_{1,i}\left(k_2,k_1\right)\right\|_{\Phi}^{2}\right)^q \le C_q\left|\frac{k_2-k_1}{\delta_{n,T}^2}\right|^q,$$

where we assume without loss of generality that $k_{n,T} < T-k_{n,T}$, so that $k_1^q\left(T-k_1\right)^{-q} < 1$. Turning to $\chi_{2,i}\left(k_2,k_1\right)$, first note that we can decompose

$$\frac{1}{n}\sum_{j=1}^{n}\left(\frac{1}{T-k_2}\underline{\mathbf{X}}_j^{(k_2)\prime}\underline{\mathbf{M}}_{T-k_2}\underline{\mathbf{u}}_j^{(k_2)} - \frac{1}{T-k_1}\underline{\mathbf{X}}_j^{(k_1)\prime}\mathbf{M}_{T-k_1}\mathbf{u}_j^{(k_1)}\right) \tag{B.28}$$
$$= \frac{1}{n(T-k_2)}\sum_{j=1}^{n}\left(\underline{\mathbf{X}}_j^{(k_2)} - \underline{\mathbf{X}}_j^{(k_1)}\right)'\underline{\mathbf{M}}_{T-k_2}\underline{\mathbf{u}}_j^{(k_2)} + \frac{1}{n(T-k_2)}\sum_{j=1}^{n}\underline{\mathbf{X}}_i^{(k_1)\prime}\left(\underline{\mathbf{M}}_{T-k_2} - \mathbf{M}_{T-k_1}\right)\underline{\mathbf{u}}_j^{(k_2)} +$$
$$\frac{1}{n(T-k_2)}\sum_{j=1}^{n}\underline{\mathbf{X}}_j^{(k_1)\prime}\left(\underline{\mathbf{M}}_{T-k_2} - \mathbf{M}_{T-k_1}\right)\left(\underline{\mathbf{u}}_j^{(k_2)} - \mathbf{u}_j^{(k_1)}\right) + \frac{k_2-k_1}{n(T-k_2)(T-k_1)}\sum_{j=1}^{n}\underline{\mathbf{X}}_j^{(k_1)\prime}\mathbf{M}_{T-k_1}\mathbf{u}_j^{(k_1)}.$$

We bound each of the four terms in this expression. Note that by Theorem B.1 and the triangle inequality, we can establish that

$$\mathrm{E}\left\|\frac{1}{n(T-k_2)}\sum_{j=1}^{n}\sum_{t=k_1}^{k_2}X_{j,t}\left(k_1\right)u_{j,t+1}'\left(k_2\right)\right\|_{\Phi}^{2q} \le C_q\mathrm{E}\left\|\frac{1}{n(T-k_2)}\sum_{j=1}^{n}\sum_{t=k_1}^{k_2}X_{j,t}\left(k_1\right)e_{j,t+1}'\right\|_{\Phi}^{2q} +$$
$$C_q\mathrm{E}\left\|\frac{1}{n(T-k_2)}\sum_{j=1}^{n}\sum_{t=k_1}^{k_2}X_{j,t}\left(k_1\right)s_{j,t}'\left(k_2\right)\right\|_{\Phi}^{2q} \le C_q\left(\frac{k_2-k_1}{n(T-k_2)^2}\right)^q\left(1+\left\|\boldsymbol{\alpha}\left(k_2\right)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{2q}\right),$$

and

$$\mathrm{E}\left\|\tfrac{1}{n(T-k_2)}\sum_{j=1}^{n}\sum_{t=k_2}^{T-1}\underline{x}_{j,t}(k_2,k_1)u'_{j,t+1}(k_2)\right\|_{\Phi}^{2q}\le C_q\mathrm{E}\left\|\tfrac{1}{n(T-k_2)}\sum_{j=1}^{n}\sum_{t=k_2}^{T-1}\underline{x}_{j,t}(k_2,k_1)e'_{j,t+1}\right\|_{\Phi}^{2q}+$$
$$C_q\mathrm{E}\left\|\tfrac{1}{n(T-k_2)}\sum_{j=1}^{n}\sum_{t=k_2}^{T-1}\underline{x}_{j,t}(k_2,k_1)s'_{j,t}(k_2)\right\|_{\Phi}^{2q}\le C_q\left(\tfrac{k_2-k_1}{n(T-k_2)}\right)^{q}\left(1+\|\boldsymbol{\alpha}(k_2)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{2q}\right),$$

noting that $\mathrm{E}\left(\underline{x}_{j,t}(k_2,k_1)s'_{j,t}(k_2)\right)=0$. Next, by Theorem B.1 and the triangle inequality, we can establish that

$$\left\|\frac{1}{n(T-k_2)}\sum_{j=1}^{n}\sum_{t=k_2}^{T-1}\mathrm{E}\left(\underline{x}_{j,t}(k_2,k_1)e'_{j,t+1}\right)\eta_t^{(T-k_2)\prime}\eta_{T-k_2+1}^{(T-k_2)}\right\|<C(T-k_2)^{-1},$$

and thus, letting $\underline{\mathbf{P}}_{T-k_2}$ be a $(T-k_1)\times(T-k_1)$ matrix with $\mathbf{P}_{T-k_2}$ in the upper left $(T-k_2)\times(T-k_2)$ block and zeros elsewhere, we have

$$\begin{aligned}\mathrm{E}&\left\|\tfrac{1}{n(T-k_2)}\sum_{j=1}^{n}\left(\underline{\mathbf{X}}_j^{(k_2)}-\underline{\mathbf{X}}_j^{(k_1)}\right)'\underline{\mathbf{P}}_{T-k_2}\underline{\mathbf{u}}_j^{(k_2)}\right\|_{\Phi}^{2q}\\&\le C_q\mathrm{E}\left\|\tfrac{1}{n(T-k_2)}\sum_{j=1}^{n}\sum_{t=k_2}^{T-1}\underline{x}_{j,t}(k_2,k_1)e'_{j,t+1}\eta_t^{(T-k_2)\prime}\eta_{T-k_2+1}^{(T-k_2)}\right\|_{\Phi}^{2q}+\\&\quad C_q\mathrm{E}\left\|\tfrac{1}{n(T-k_2)}\sum_{j=1}^{n}\sum_{t=k_2}^{T-1}\underline{x}_{j,t}(k_2,k_1)s'_{j,t}(k_2)\eta_t^{(T-k_2)\prime}\eta_{T-k_2+1}^{(T-k_2)}\right\|_{\Phi}^{2q}\\&\le C_q\left[\left(\tfrac{k_2-k_1}{(T-k_2)^2}\right)^{q}+\left(\tfrac{k_2-k_1}{n(T-k_2)^2}\right)^{q}\|\boldsymbol{\alpha}(k_2)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{2q}\right].\end{aligned}$$

Thus the first term in (B.28) satisfies

$$\mathrm{E}\left\|\tfrac{1}{n(T-k_2)}\sum_{j=1}^{n}\left(\underline{\mathbf{X}}_j^{(k_2)}-\underline{\mathbf{X}}_j^{(k_1)}\right)'\underline{\mathbf{M}}_{T-k_2}\underline{\mathbf{u}}_j^{(k_2)}\right\|_{\Phi}^{2q}\le C_q\left(\tfrac{k_2-k_1}{\delta_{n,T}^2}\right)^{q}\left(1+\|\boldsymbol{\alpha}(k_2)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{2q}\right).$$

Next, for the second term in (B.28), note that $\mathbf{P}_{T-k_1}-\underline{\mathbf{P}}_{T-k_2}=\boldsymbol{\eta}_{T-k_1}\boldsymbol{\eta}'_{T-k_1}-\underline{\boldsymbol{\eta}}_{T-k_2}\underline{\boldsymbol{\eta}}'_{T-k_2}$, where $\underline{\boldsymbol{\eta}}_{T-k_2}:=\left[\boldsymbol{\eta}'_{T-k_2}:\mathbf{0}_{p\times(k_2-k_1)}\right]'$. Then by the Cauchy–Schwarz inequality we have

$$\begin{aligned}&\left\|\mathrm{E}\left(\tfrac{1}{n(T-k_1)}\sum_{j=1}^{n}\underline{\mathbf{X}}_j^{(k_1)\prime}\left(\boldsymbol{\eta}_{T-k_1}-\underline{\boldsymbol{\eta}}_{T-k_2}\right)\underline{\boldsymbol{\eta}}'_{T-k_2}\underline{\mathbf{u}}_j^{(k_2)}\right)\right\|_{\Phi}^{2}\le\left\|\boldsymbol{\eta}_{T-k_1}-\underline{\boldsymbol{\eta}}_{T-k_2}\right\|^{2}\left\|\boldsymbol{\eta}_{T-k_2}\right\|^{2}\times\\&\mathrm{tr}\left(\mathrm{E}\left(\tfrac{1}{n(T-k_1)}\sum_{j=1}^{n}\underline{\mathbf{X}}_j^{(k_1)\prime}\underline{\mathbf{X}}_j^{(k_1)}\right)\right)\cdot\mathrm{tr}\left(\mathrm{E}\left(\tfrac{1}{n(T-k_1)}\sum_{j=1}^{n}\underline{\mathbf{u}}_j^{(k_2)\prime}\underline{\mathbf{u}}_j^{(k_2)}\right)\Phi\right)\le C\tfrac{(k_2-k_1)k_1}{(T-k_2)^2}\tfrac{1+\|\boldsymbol{\alpha}(k_2)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{2}}{T-k_2}\end{aligned}$$

and

$$\left\|\mathrm{E}\left(\tfrac{1}{n(T-k_1)}\sum_{j=1}^{n}\underline{\mathbf{X}}_j^{(k_1)\prime}\boldsymbol{\eta}_{T-k_1}\left(\boldsymbol{\eta}_{T-k_1}-\underline{\boldsymbol{\eta}}_{T-k_2}\right)'\underline{\mathbf{u}}_j^{(k_2)}\right)\right\|_{\Phi}^{2}\le C\tfrac{k_2-k_1}{(T-k_2)^2}\tfrac{k_1}{T-k_1}\left(1+\|\boldsymbol{\alpha}(k_2)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{2}\right).$$

Using these results in conjunction with Theorem B.1 we have

$$\mathrm{E}\left\|\frac{1}{n(T-k_1)}\sum_{j=1}^{n}\underline{\mathbf{X}}_j^{(k_1)\prime}\left(\mathbf{P}_{T-k_1}-\underline{\mathbf{P}}_{T-k_2}\right)\underline{\mathbf{u}}_j^{(k_2)}\right\|_{\Phi}^{2q}\le C(k_2-k_1)^{q}\delta_{n,T}^{-2q},$$

where we assume without loss of generality that $k_2 - k_1 < T - k_2$. Using the result established above and the triangle and Cauchy–Schwarz inequalities, the second term in (B.28) satisfies

$$\mathrm{E}\left\| \frac{1}{n(T-k_1)} \sum_{j=1}^{n} \underline{\mathbf{X}}_j^{(k_1)\prime} \left(\underline{\mathbf{M}}_{T-k_2} - \mathbf{M}_{T-k_1} \right) \underline{\mathbf{u}}_j^{(k_2)} \right\|_{\Phi}^{2q} \leq C_q \mathrm{E} \left\| \frac{1}{n(T-k_1)} \sum_{j=1}^{n} \sum_{t=k_1}^{k_2} X_{j,t}(k_1) u'_{j,t+1}(k_2) \right\|_{\Phi}^{2q} + C_q \mathrm{E} \left\| \frac{1}{n(T-k_1)} \sum_{j=1}^{n} \underline{\mathbf{X}}_j^{(k_1)\prime} \left(\mathbf{P}_{T-k_1} - \underline{\mathbf{P}}_{T-k_2} \right) \underline{\mathbf{u}}_j^{(k_2)} \right\|_{\Phi}^{2q} \leq C (k_2 - k_1)^q \delta_{n,T}^{-2q} .$$

Next, for the third term in (B.28), it is straightforward to establish that

$$\left\| \sum_{t=k_1}^{T-1} \sum_{s=k_1}^{k_2-1} \mathrm{E}\left(X_{j,t}(k_1) e'_{j,s+1} \right) \eta_t^{(T-k_1)\prime} \eta_s^{(T-k_1)} \right\| < C \left((T-k_1)^{-1} (k_2-k_1)^{\frac{1}{2}} \right),$$

and

$$\left\| \sum_{t=k_1}^{T-1} \sum_{s=k_1}^{k_2-1} \mathrm{E}\left(X_{j,t}(k_1) s'_{j,s}(k_2) \right) \eta_t^{(T-k_1)\prime} \eta_s^{(T-k_1)} \right\| < C \left((T-k_1)^{-1} (k_2-k_1)^{\frac{1}{2}} \|\boldsymbol{\alpha}(k_2) - \boldsymbol{\alpha}\|_{\Gamma\otimes\Phi} \right),$$

and so using these results in conjunction with Theorem B.1, the third term in (B.28) satisfies

$$\mathrm{E}\left\| \frac{1}{n(T-k_1)} \sum_{j=1}^{n} \underline{\mathbf{X}}_j^{(k_1)\prime} \mathbf{M}_{T-k_1} \left(\underline{\mathbf{u}}_j^{(k_2)} - \mathbf{u}_j^{(k_1)} \right) \right\|_{\Phi}^{2q} \leq C_q \mathrm{E} \left\| \sum_{j=1}^{n} \sum_{t=k_1}^{T-1} \frac{1}{n(T-k_1)} X_{j,t}(k_1) u'_{j,t+1}(k_2) \right\|_{\Phi}^{2q} + C_q \mathrm{E} \left\| \frac{1}{n(T-k_1)} \sum_{j=1}^{n} \sum_{t=k_1}^{T-1} \sum_{s=k_1}^{k_2-1} X_{j,t}(k_1) \eta_t^{(T-k_1)\prime} \eta_s^{(T-k_1)} u'_{j,s+1}(k_2) \right\|_{\Phi}^{2q} \leq C_q \left(\frac{k_2-k_1}{\delta_{n,T}^2} \right)^q .$$

Finally, by Lemmas B.3(iii) and (ix), the final term in (B.28) satisfies

$$\mathrm{E}\left\| \frac{k_2-k_1}{n(T-k_2)(T-k_1)} \sum_{j=1}^{n} \underline{\mathbf{X}}_j^{(k_1)\prime} \mathbf{M}_{T-k_1} \mathbf{u}_j^{(k_1)} \right\|_{\Phi}^{2q} \leq C_q \frac{(k_2-k_1)^{2q}}{(T-k_1)^{2q}} \frac{k_1^q}{\delta_{n,T-k_2}^{2q}} \leq C_q \left| \frac{k_2-k_1}{\delta_{n,T}^2} \right|^q ,$$

where we assume that $k_2 - k_1 < \delta_{n,T}$ without loss of generality. Thus

$$\mathrm{E}\left\| \frac{1}{n} \sum_{j=1}^{n} \left(\frac{1}{T-k_2} \underline{\mathbf{X}}_j^{(k_2)\prime} \underline{\mathbf{M}}_{T-k_2} \underline{\mathbf{u}}_j^{(k_2)} - \frac{1}{T-k_1} \underline{\mathbf{X}}_j^{(k_1)\prime} \mathbf{M}_{T-k_1} \mathbf{u}_j^{(k_1)} \right) \right\|_{\Phi}^{2q} \leq C_q \left| \frac{k_2-k_1}{\delta_{n,T}^2} \right|^{2q} .$$

Together with the triangle and Hölder's inequalities, these results imply that

$$\mathrm{E}\left(\frac{1}{n} \sum_{i=1}^{n} \|\chi_{2,i}(k_1,k_2)\|_{\Phi}^2 \right)^q \leq C_q \frac{(k_2-k_1)^q}{\delta_{n,T}^{2q}} .$$

Next we consider $\chi_{3,i}(k_1,k_2)$. Note that we can decompose $\chi_{3,i}(k_1,k_2)$ as follows

$$\chi_{3,i}(k_1,k_2) = \ddot{X}'_{i,T}(k_1) \hat{\mathbf{Q}}_{(k_1)}^{-1} \boldsymbol{\Delta}_{k_2,k_1} \mathbf{J}'_{k_2,k_1} \hat{\mathbf{Q}}_{(k_2)}^{-1} \frac{1}{n(T-k_2)} \sum_{j=1}^{n} \underline{\mathbf{X}}_j^{(k_2)\prime} \underline{\mathbf{M}}_{T-k_2} \underline{\mathbf{u}}_j^{(k_2)},$$

where $\boldsymbol{\Delta}_{k_2,k_1}$ is a $k_1 \times k_2$ matrix defined as

$$\mathbf{\Delta}_{k_2,k_1} := \tfrac{1}{n(T-k_1)} \sum_{j=1}^{n} \sum_{t=k_1}^{T-1} X_{j,t}(k_1)\underline{X}'_{j,t}(k_1) - \tfrac{1}{n(T-k_2)} \sum_{j=1}^{n} \sum_{t=k_2}^{T-1} X_{j,t}(k_1) X'_{j,t}(k_2) + \tfrac{1}{n(T-k_2)} \sum_{j=1}^{n} \mathbf{X}_j^{(k_1)\prime} \left(\mathbf{P}_{T-k_1} - \underline{\mathbf{P}}_{T-k_2}\right) \underline{\mathbf{X}}_j^{(k_1)}.$$

Now using similar arguments as above we have

$$\mathrm{E}\left\|\sum_{j=1}^{n} \mathbf{X}_j^{(k_1)\prime} \left(\mathbf{P}_{T-k_1} - \underline{\mathbf{P}}_{T-k_2}\right) \underline{\mathbf{X}}_j^{(k_1)}\right\|_{\Phi}^{2q} \le \mathrm{E}\left\|\sum_{j=1}^{n} \underline{\mathbf{X}}_j^{(k_1)\prime} \left(\mathbf{P}_{T-k_1} - \underline{\mathbf{P}}_{T-k_2}\right) \underline{\mathbf{X}}_j^{(k_1)}\right\|_{\Phi}^{2q} \le C_q \left(\tfrac{k_2-k_1}{(T-k_2)^2}\right)^q,$$

while we can decompose

$$\begin{aligned} &\tfrac{1}{n(T-k_1)} \sum_{j=1}^{n} \sum_{t=k_1}^{T-1} X_{j,t}(k_1)\underline{X}'_{j,t}(k_1) - \tfrac{1}{n(T-k_2)} \sum_{j=1}^{n} \sum_{t=k_2}^{T-1} X_{j,t}(k_1) X'_{j,t}(k_2) \\ &= \tfrac{1}{n(T-k_1)} \sum_{j=1}^{n} \sum_{t=k_1}^{k_2-1} X_{j,t}(k_1)\underline{X}'_{j,t}(k_1) + \tfrac{k_2-k_1}{n(T-k_1)(T-k_2)} \sum_{j=1}^{n} \sum_{t=k_2+1}^{T-1} X_{j,t}(k_1)\underline{X}'_{j,t}(k_1) + \\ &\tfrac{1}{n(T-k_2)} \sum_{j=1}^{n} \sum_{t=k_2}^{T-1} \left(X_{j,t}(k_1)\underline{X}'_{j,t}(k_1) - X_{j,t}(k_1) X'_{j,t}(k_2)\right). \end{aligned}$$

By Theorem B.1 we have

$$\mathrm{E}\left\|\frac{1}{n(T-k_1)} \sum_{j=1}^{n} \sum_{t=k_1}^{k_2-1} X_{j,t}(k_1)\underline{X}'_{j,t}(k_1)\right\|^{2q} \le C_q \left(\frac{(k_2-k_1)^q}{n^q(T-k_1)^{2q}} + \frac{1}{(T-k_1)^{2q}}\right),$$

$$\mathrm{E}\left\|\frac{k_2-k_1}{T-k_1}\frac{1}{n(T-k_2)} \sum_{j=1}^{n} \sum_{t=k_2}^{T-1} X_{j,t}(k_1)\underline{X}'_{j,t}(k_1)\right\|^{2q} \le C_q \left(\frac{(k_2-k_1)^{2q}}{(T-k_1)^{2q}} + \frac{(k_2-k_1)^{2q} k_1^q}{n^q(T-k_1)^{3q}}\right),$$

and

$$\mathrm{E}\left\|\tfrac{1}{n(T-k_2)} \sum_{j=1}^{n} \sum_{t=k_2}^{T-1} \left(X_{j,t}(k_1)\underline{X}'_{j,t}(k_1) - X_{j,t}(k_1) X'_{j,t}(k_2)\right) - \underline{\Gamma}_{(k_1,k_2)}\right\|^{2q} \le C_q \tfrac{(k_2-k_1)^q}{n^q(T-k_2)^q},$$

where $\underline{\Gamma}_{(k_1,k_2)} := \mathrm{E}\left[\frac{1}{n}\sum_j \left(X_{j,t}(k_1)\underline{X}'_{j,t}(k_1) - X_{j,t}(k_2) X'_{j,t}(k_2)\right)\right]$ can be partitioned as follows:

$$\underline{\Gamma}_{(k_1,k_2)} = -\left[\mathbf{0}_{mk_1 \times m(k_2-k_1)} : \mathrm{E}\left(\tfrac{1}{n}\sum_j X_{j,t}(k_1)\underline{x}'_{j,t}(k_2,k_1)\right)\right] = -\left[\mathbf{0}_{mk_1 \times m(k_2-k_1)} : \boldsymbol{\lambda}_{k_2,k_1}\right]$$

for suitably defined $\boldsymbol{\lambda}_{k_2,k_1}$. These results together with Lemmas B.3(xi) and (xii) imply that

$$\mathrm{E}\left\|\hat{\mathbf{Q}}_{(k_1)}^{-1}\mathbf{J}'_{k_2,k_1} - \mathbf{J}'_{k_2,k_1}\hat{\mathbf{Q}}_{(k_2)}^{-1} - \Gamma^{-1}(k_1)\boldsymbol{\lambda}_{k_2,k_1}\mathbf{J}'_{k_2,k_2-k_1}\Gamma^{-1}(k_2)\right\|_{\Phi}^{q} \le C_q \frac{(k_2-k_1)^q}{\delta_{n,T}^{2q}}.$$

Then using the same arguments used to establish (B.27) we can show that

$$\mathrm{E}\left\|\frac{1}{n(T-k_2)}\sum_j \mathbf{J}'_{k_2,k_2-k_1}\Gamma^{-1}(k_2)\underline{\mathbf{X}}_j^{(k_2)\prime}\underline{\mathbf{M}}_{T-k_2}\underline{\mathbf{u}}_j^{(k_2)}\right\|_{\Phi}^{q} \le C_q \frac{(k_2-k_1)^q}{\delta_{n,T}^{2q}}.$$

These results, together with the triangle and Hölder's inequalities, imply that

$$\mathrm{E}\left(\frac{1}{n}\left\|\sum_{i=1}^{n}\chi_{3,i}(k_1,k_2)\right\|_{\Phi}^{2}\right)^q \le C_q \frac{(k_2-k_1)^q}{\delta_{n,T}^{2q}}.$$

Next we consider $\chi_{4,i}(k_2,k_1)$. Recall that $\mathbf{e}_i^{(k_1)} = \left[e_{i,k_1+1} : \cdots : e_{i,T}\right]'$. Since $\chi_{4,i}(k_1,k_2) = \bar{e}_{i,T}(k_2) - \bar{e}_{i,T}(k_1) = \mathbf{e}_i^{(k_1)\prime}\boldsymbol{\eta}_{\boldsymbol{\Delta}}$, by Theorem B.1 it follows that

$$\mathrm{E}\left(\frac{1}{n}\left\|\sum_{i=1}^{n}\chi_{4,i}(k_1,k_2)\right\|_{\Phi}^{2}\right)^{q} = \mathrm{E}\left(\frac{1}{n}\left\|\sum_{i=1}^{n}\mathbf{e}_i^{(k_1)\prime}\boldsymbol{\eta}_{\boldsymbol{\Delta}}\right\|_{\Phi}^{2}\right)^{q} \leq C_q (k_2-k_1)^q (T-k_1)^{-2q}.$$

Finally, turning to $\chi_{5,i}(k_2,k_1) = \bar{s}_{i,T}(k_2) - \bar{s}_{i,T}(k_1) - \left(s_{i,T}(k_2) - s_{i,T}(k_1)\right)$, note that

$$\begin{aligned}\mathrm{E}\left(\frac{1}{n}\left\|\sum_{i=1}^{n}s_{i,T}(k_2) - s_{i,T}(k_1)\right\|_{\Phi}^{2}\right)^{q} &\leq C_q \left\|\boldsymbol{\alpha}(k_2) - \boldsymbol{\alpha}(k_1)\right\|_{\Gamma\otimes\Phi}^{2q} \\ &\leq C_q\left(\left\|\boldsymbol{\alpha}(k_2)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{2q} - \left\|\boldsymbol{\alpha}(k_1)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{2q}\right) \\ &\leq C_q\left(\left|l^q(k_2) - l^q(k_1)\right| + \delta_{n,T}^{-2q}\left|k_2-k_1\right|^q\right).\end{aligned}$$

Next, note that we can decompose

$$\bar{s}_{i,T}(k_2) - \bar{s}_{i,T}(k_1) = \mathbf{s}_i^{(k_1)\prime}\boldsymbol{\eta}_{\boldsymbol{\Delta}} + \left(\mathbf{s}_i^{(k_2)} - \mathbf{J}_{T-k_2,T-k_1}\mathbf{s}_i^{(k_1)}\right)'\eta_{T-k_2}\eta_{T-k_2+1}^{(T-k_2)}.$$

Then by the triangle inequality we can establish that

$$\begin{aligned}\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\boldsymbol{\eta}_{\boldsymbol{\Delta}}'\mathbf{s}_i^{(k_1)}\right\|_{\Phi}^{2}\right)^{q} &\leq C_q\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{s,t=k_2-k_1+1}^{T-1}\mathrm{tr}\left[\eta_{\boldsymbol{\Delta},t}^{(T,k_2,k_1)}\eta_{\boldsymbol{\Delta},s}^{(T,k_2,k_1)}\left(s_{i,t}(k_1)s_{i,s}'(k_1)\right)\Phi\right]\right)^{q} + \\ &C_q\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\sum_{s,t=1}^{k_2-k_1}\mathrm{tr}\left[\eta_{\boldsymbol{\Delta},t}^{(T,k_2,k_1)}\eta_{\boldsymbol{\Delta},s}^{(T,k_2,k_1)}\left(s_{i,t}(k_1)s_{i,s}'(k_1)\right)\Phi\right]\right) \leq C_q\frac{(k_2-k_1)^q}{(T-k_1)^{2q}}\left\|\boldsymbol{\alpha}(k_1)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{2q},\end{aligned}$$

where the final inequality follows by application of Theorem B.1. Meanwhile, by the Cauchy–Schwarz inequality,

$$\begin{aligned}\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\left(\mathbf{s}_i^{(k_2)} - \mathbf{J}_{T-k_2,T-k_1}\mathbf{s}_i^{(k_1)}\right)'\boldsymbol{\eta}_{T-k_2}\eta_{T-k_2+1}^{(T-k_2)}\right\|_{\Phi}^{2}\right)^{q} &\leq C_q\frac{\left\|\boldsymbol{\alpha}(k_2)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{2q} - \left\|\boldsymbol{\alpha}(k_1)-\boldsymbol{\alpha}\right\|_{\Gamma\otimes\Phi}^{2q}}{(T-k_1)^q} \\ &\leq C_q\frac{1}{(T-k_1)^q}\left(\left|l^q(k_2)-l^q(k_1)\right| + \delta_{n,T}^{-2q}(k_2-k_1)^q\right).\end{aligned}$$

Therefore, by the triangle inequality we have

$$\mathrm{E}\left(\frac{1}{n}\left\|\sum_{i=1}^{n}\chi_{5,i}(k_1,k_2)\right\|_{\Phi}^{2}\right)^{q} \leq C_q\left((k_2-k_1)^q\delta_{n,T}^{-2q} + l^q(k_2) - l^q(k_1)\right),$$

completing the proof. ■

For notational brevity, let $\vartheta(k)$ denote the expected second order QFR evaluated at k, i.e.:

$$\vartheta(k) := \mathrm{E}\left(\mathrm{tr}\left(\Phi\left(\frac{1}{n}\sum_{i=1}^{n}\hat{u}_{i,T+1}(k,k)\hat{u}_{i,T+1}'(k,k) - \left(1+\frac{p^2}{T}\right)\Sigma\right)\right)\right).$$

Similarly, let $k^* := k_{n,T}^*$ and $\hat{u}_{i,T+1}(k) := \hat{u}_{i,T+1}(k,k)$ throughout this subsection for brevity. The proof demonstrates that $\mathrm{E}\left[l^{-1}(k^*)\vartheta\left(\hat{k}\right)\right]$ has a limit of 1. We begin by decomposing

$$\mathrm{E}\left(\frac{\vartheta\left(\hat{k}\right)}{l(k^*)}\right) = \mathrm{E}\left(\frac{\vartheta(k^*)}{l(k^*)}\right) + \mathrm{E}\left(\frac{\vartheta\left(\hat{k}\right) - \vartheta(k^*)}{l(k^*)}\right). \tag{B.29}$$

The first term has a limit of 1 under Theorem 3.1. Thus we need to show that the second term is $o(1)$. Now by Hölder's inequality, for all $r > 1$ we have

$$\mathrm{E}\left(\frac{\vartheta\left(\hat{k}\right)-\vartheta\left(k^*\right)}{l(k^*)}\right) \leq \left\{\mathrm{E}\left(\frac{l\left(\hat{k}\right)}{l(k^*)}\right)^r\right\}^{\frac{1}{r}} \left\{\mathrm{E}\left(\frac{\vartheta\left(\hat{k}\right)-\vartheta\left(k^*\right)}{l\left(\hat{k}\right)}\right)^{\frac{r}{r-1}}\right\}^{\frac{r-1}{r}}. \tag{B.30}$$

We show that the expectation in the first set of braces on the right hand side of the above has a limit of one by showing that

$$\lim_{n,T\to\infty} \mathrm{E}\left(\frac{l\left(\hat{k}\right)}{l(k^*)}-1\right)^r = 0 \tag{B.31}$$

for all $r \geq 1$, which implies that $\lim_{n,T\to\infty} \mathrm{E}\left|\frac{l\left(\hat{k}\right)}{l(k^*)}\right|^r = 1$. The proof of this result relies on defining the following set.

$$A_\epsilon := \left\{k : \frac{l(k)}{l(k^*)} - 1 > \epsilon_0, 1 \leq k \leq k_{n,T}\right\}$$

for all $\epsilon_0 > 0$. Then

$$\begin{aligned}
\mathrm{E}\left(\frac{l\left(\hat{k}\right)}{l(k^*)}-1\right)^r &= \sum_{k=1}^{k_{n,T}} \left(\frac{l(k)}{l(k^*)}-1\right)^r \mathrm{P}\left(\hat{k}=k\right)\\
&\leq \epsilon_0^r + \sum_{k\in A_\epsilon} \left(\frac{l(k)}{l(k^*)}-1\right)^r \mathrm{P}\left(\hat{l}(k) \leq \hat{l}\left(k^*\right)\right),
\end{aligned}$$

so that (B.31) follows if the second term above is $o(1)$. We first bound $\mathrm{P}\left(\hat{l}(k) \leq \hat{l}\left(k^*\right)\right)$ as follows. For all $q = 1, 2, \ldots,$ we have

$$\begin{aligned}
\mathrm{P}\left(\hat{l}(k) \leq \hat{l}\left(k^*\right)\right) &= \mathrm{P}\left(\frac{\hat{l}(k)}{l(k)} \leq \frac{\hat{l}(k^*)}{l(k)}\right) = \mathrm{P}\left(\frac{\hat{l}(k)-l(k)-\hat{l}(k^*)+l(k^*)}{l(k)} \leq \frac{l(k^*)-l(k)}{l(k)}\right)\\
&\leq \mathrm{P}\left(\frac{\left|\hat{l}(k^*)-l(k^*)-\hat{l}(k)+l(k)\right|}{l(k)} \geq \frac{l(k)-l(k^*)}{l(k)}\right)\\
&\leq \left(\frac{l(k)-l(k^*)}{l(k)}\right)^{-q} \mathrm{E}\left(\left|\frac{\hat{l}(k^*)-l(k^*)-\hat{l}(k)+l(k)}{l(k)}\right|^q\right),
\end{aligned} \tag{B.32}$$

where the second inequality follows from Markov's inequality. Substituting in this bound we have

$$\begin{aligned}
\sum_{k\in A_\epsilon} \left(\frac{l(k)}{l(k^*)}-1\right)^r \mathrm{P}\left(\hat{l}(k) \leq \hat{l}\left(k^*\right)\right) &\leq \sum_{k\in A_\epsilon} \left(\frac{l(k)-l(k^*)}{l(k^*)}\right)^r \left(\frac{l(k)-l(k^*)}{l(k)}\right)^{-q} \frac{\mathrm{E}\left(\left|\hat{l}(k^*)-l(k^*)-\hat{l}(k)+l(k)\right|^q\right)}{l^q(k)}\\
&= \sum_{k\in A_\epsilon} \left(\frac{l(k)}{l(k)-l(k^*)}\right)^{q-r} \left(\frac{l(k)}{l(k^*)}\right)^r \frac{\mathrm{E}\left(\left|\hat{l}(k^*)-l(k^*)-\hat{l}(k)+l(k)\right|^q\right)}{l^q(k)}\\
&\leq \left(\frac{1+\epsilon_0}{\epsilon_0}\right)^{q-r} \sum_{k\in A_\epsilon} \left(\frac{l(k)}{l(k^*)}\right)^r \frac{\mathrm{E}\left(\left|\hat{l}(k^*)-l(k^*)-\hat{l}(k)+l(k)\right|^q\right)}{l^q(k)}
\end{aligned}$$

$$\leq C_q \left(\frac{1+\epsilon_0}{\epsilon_0}\right)^{q-r} \left(\frac{1}{k^*}\right)^r \left\{\sum_{k=1}^{k_{n,T}} \delta_{n,T-k^*}^{2r} l^r(k) \frac{\mathrm{E}\left(\left|\hat{l}(k^*)-l(k^*)-\hat{l}(k)+l(k)\right|^q\right)}{l^q(k)}\right\}$$
$$= o(1)$$

for any $q \geq 1$. The term in the braces is $o(1)$ by Lemma B.8(ii) together with the fact that $\delta_{n,T-k^*}^{2r} = \delta_{n,T}^{2r} + o(1)$.

The remainder of the proof shows that the expectation in the second set of braces on the right hand side of (B.30) is $o(1)$. It is sufficient to establish that

$$\mathrm{E}\left\{\left(l^{-1}\left(\hat{k}\right)\frac{1}{n}\sum_{i=1}^{n}\left\|\hat{u}_{i,T+1}\left(\hat{k}\right)-\hat{u}_{i,T+1}\left(k^*\right)\right\|_{\Phi}^{2}\right)^{q}\right\} = o(1) \tag{B.33}$$

for some $q > 1$. Now for any $r > 1$ we have

$$\mathrm{E}\left[\frac{\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\hat{u}_{i,T+1}\left(\hat{k}\right)-\hat{u}_{i,T+1}\left(k^*\right)\right\|_{\Phi}^{2}\right)^{q}}{l^q\left(\hat{k}\right)}\right] \leq \sum_{k=1}^{k_{n,T}}\left\{\frac{\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\hat{u}_{i,T+1}\left(\hat{k}\right)-\hat{u}_{i,T+1}\left(k^*\right)\right\|_{\Phi}^{2}\right)^{qr}}{l^{qr}(k)}\right\}^{\frac{1}{r}}\left\{P\left(\hat{k}=k\right)\right\}^{\frac{r-1}{r}}.$$

Now using Lemma B.9 we have

$$\sum_{k=1}^{k_{n,T}}\left\{l^{-qr}(k)\,\mathrm{E}\left(\frac{1}{n}\sum_{i=1}^{n}\left\|\hat{u}_{i,T+1}\left(\hat{k}\right)-\hat{u}_{i,T+1}\left(k^*\right)\right\|_{\Phi}^{2}\right)^{qr}\right\}^{\frac{1}{r}}\left\{P\left(\hat{k}=k\right)\right\}^{\frac{r-1}{r}} \tag{B.34}$$
$$\leq C_q\sum_{k=1}^{k_{n,T}}\left(\frac{|k-k^*|}{\delta_{n,T}^2 l(k)}\right)^q\left\{P\left(\hat{k}=k\right)\right\}^{\frac{r-1}{r}} + C_q\sum_{k=1}^{k_{n,T}}\left(\frac{l(k)-l(k^*)}{l(k)}\right)^q\left\{P\left(\hat{k}=k\right)\right\}^{\frac{r-1}{r}}.$$

Using Assumption 4 and the bound for $P\left(\hat{k}=k\right)$ given in (B.32), the first term on the right hand side of the above is bounded by

$$k^{*(\pi-1)q} + \sum_{k\in A_{\pi,n,T}}\left(\frac{|k-k^*|}{\delta_{n,T}^2 l(k)}\right)^q\left(\frac{l(k)}{l(k)-l(k^*)}\right)^q \mathrm{E}\left(\left|\frac{\hat{l}(k^*)-l(k^*)-\hat{l}(k)+l(k)}{l(k)}\right|^q\right)$$
$$\leq k^{*(\pi-1)q} + C_q k^{*\xi q}\sum_{k=1}^{k_{n,T}}\frac{\mathrm{E}\left(\left|\hat{l}(k)-l(k)-\hat{l}(k^*)+l(k^*)\right|^q\right)}{l^q(k)},$$

where $A_{\pi,n,T} := \left\{k : \left|k-k_{n,T}^*\right| > \left(k_{n,T}^*\right)^{\pi}\right\}$. The first term in the above expression is $o(1)$. Turning to the second term, by Lemma B.8(i) we can substitute in the following bound

$$\frac{\mathrm{E}\left(\left|\hat{l}(k)-l(k)-\hat{l}(k^*)+l(k^*)\right|^q\right)}{l^q(k)} \leq C_q\left(\frac{k^q+k^{*q}}{l^q(k)\delta_{n,T}^{2q}}\frac{1}{\left(T-k_{n,T}\right)^{q/2}} + \frac{k^q+k^{*q}}{l^q(k)n^{q/2}\left(T-k_{n,T}\right)^{3q/2}}\right.$$
$$\left. + \frac{\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{2q}+\|\boldsymbol{\alpha}(k^*)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{2q}}{\delta_{n,T}^{q}l^q(k)} + \frac{\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{q}+\|\boldsymbol{\alpha}(k^*)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{q}}{\delta_{n,T}^{q}l^q(k)}\right),$$

which yields

$$k^{*\xi q}\sum_{k=1}^{k_{n,T}}\frac{\mathrm{E}\left(\left|\hat{l}(k)-l(k)-\hat{l}(k^*)+l(k^*)\right|^q\right)}{l^q(k)}$$

$$\leq C_q k^{*\xi q}\sum_{k=1}^{k_{n,T}}\left(\frac{k^q+k^{*q}}{l^q(k)\delta_{n,T}^{2q}}\frac{1}{(T-k_{n,T})^{q/2}}+\frac{k^q+k^{*q}}{l^q(k)\delta_{n,T}^{2q}}\frac{\delta_{n,T}^{2q}}{n^{q/2}(T-k_{n,T})^{3q/2}}+\right.$$
$$\left.\frac{\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{2q}+\|\boldsymbol{\alpha}(k^*)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{2q}}{\delta_{n,T}^{q}l^q(k)}+\frac{\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{q}+\|\boldsymbol{\alpha}(k^*)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{q}}{\delta_{n,T}^{q}l^q(k)}\right)$$

$$\leq C_q k^{*\xi q}\sum_{k=1}^{k_{n,T}}\left(\frac{1}{(T-k_{n,T})^{q/2}}+\frac{\delta_{n,T}^{2q}}{n^{q/2}(T-k_{n,T})^{3q/2}}\right)+C_q k^{*\xi q}\left(\sum_{k=1}^{k^*}k^{*-\frac{q}{2}}+k^{*-\frac{q}{2}\underline{\pi}}\sum_{k=k^*+1}^{k_{n,T}}k^{-\frac{q}{2}(1-\underline{\pi})}\right)$$

$$\leq C_q\left(\frac{k^{*\xi q}k_{n,T}}{(T-k_{n,T})^{q/2}}+\frac{k^{*\xi q}k_{n,T}\delta_{n,T}^{2q}}{n^{q/2}(T-k_{n,T})^{3q/2}}+k^{*\xi q+1-\frac{q}{2}}+k^{*\xi q-\frac{q}{2}\underline{\pi}}\sum\nolimits_{k=1}^{k_{n,T}}k^{-\frac{q}{2}(1-\underline{\pi})}\right)$$

for any $\underline{\pi}\in(0,1)$. We then select ξ, q and $\underline{\pi}$ to ensure that the above is $o(1)$. The first term is $o(1)$ if $\xi<1-\frac{1}{q}$. For the second term to be $o(1)$ we require $\xi<\upsilon$ and $(\upsilon-\xi)^{-1}<q$. For the third term to be $o(1)$ we require $\xi<\frac{1}{2}-\frac{1}{q}$. For the final term we require $\xi<\frac{1}{2}\underline{\pi}$ for $k^{*\xi q-\frac{q}{2}\underline{\pi}}=o(1)$ and $\underline{\pi}<1-\frac{2}{q}$ for $\sum_{k=1}^{k_{n,T}}k^{-\frac{q}{2}(1-\underline{\pi})}=O(1)$, which in turn requires $q>2$ to ensure that $0<\underline{\pi}<1$. Both $\xi<\frac{1}{2}\underline{\pi}$ and $\underline{\pi}<1-\frac{2}{q}$ are satisfied if $\xi<\frac{1}{2}-\frac{1}{q}$. Thus if we set $q>2$ and choose ξ such that $\xi<\frac{1}{2}-\frac{1}{q}$ and $\xi<\upsilon$ the entire term above is $o(1)$.

Returning to the second term on the right hand side of (B.34), by using Assumption 4 and the bound for $P\left(\hat{k}=k\right)$ given in (B.32) we have

$$\sum_{k=1}^{k_{n,T}}\left(\frac{l(k)-l(k^*)}{l(k)}\right)^q\left\{P\left(\hat{k}=k\right)\right\}^{\frac{r-1}{r}}\leq\sum_{k=1}^{k_{n,T}}\frac{\mathrm{E}\left(\left|\hat{l}(k)-l(k)-\hat{l}(k^*)+l(k^*)\right|^q\right)}{l^q(k)}=o(1)$$

using the same arguments as above. Thus we have established that (B.33) holds, and this implies that

$$l^{-q}\left(\hat{k}\right)\mathrm{E}\left(\vartheta\left(\hat{k}\right)-\vartheta\left(k^*\right)\right)^q=o(1),$$

which completes the proof.

B.3. Proofs of Bias-Corrected Least Squares Theorems

Sections B.3.1 and B.3.3 below provide the proof to Theorems 3.3 and 3.4, respectively. In Section B.3.2 we establish that the QFR estimator is asymptotically unbiased in an analogous result to Theorem B.4. The following decomposition of the BCLS residual proves useful throughout.

$$\tilde{u}_{i,t+1}(k,l)=\frac{1}{n(T-l)}\sum\nolimits_{j=1}^{n}\left(\mathbf{X}_j^{(k,l)\prime}\mathbf{M}_{T-l}\mathbf{u}_j^{(k)}+n\hat{\boldsymbol{\xi}}_{(k,l)}\right)'\hat{\mathbf{Q}}_{(k,l)}^{-1}\ddot{X}_{i,t}(k,l)-\ddot{u}_{i,t+1}(k,l)$$
$$=\hat{u}_{i,t+1}(k,l)+\frac{1}{T-l}\hat{\boldsymbol{\xi}}_{(k,l)}'\hat{\mathbf{Q}}_{(k,l)}^{-1}\ddot{X}_{i,t}(k,l).\qquad\textbf{(B.35)}$$

B.3.1. Proof of Theorem 3.3 (*Asymptotic QFR*). The proof to Theorem 3.3 follows many of the same steps as outlined in the proof to Theorem 3.1. We outline the key differences here. First, using the bound given in (B.17) alongside the same arguments used to establish (B.11), we have

$$\left|\mathrm{E}\left\|\ddot{\mathbf{Q}}_T^{\frac{1}{2}}\hat{\mathbf{Q}}_{(k,k)}^{-1}\left(\frac{1}{n(T-k)}\sum_{i=1}^{n}\mathbf{X}_i^{(k,k)\prime}\mathbf{M}\mathbf{e}_i^{(k)}+\frac{p}{T-k}\hat{\boldsymbol{\xi}}_{(k,k)}\right)\right\|_\Phi^2-\frac{k\mathrm{tr}(\Phi\Pi(k))}{n\left(T-k_{n,T}\right)}\right|$$
$$\leq C\left(\frac{k}{\delta_{n,T-k}^2(T-k)^{1/2}}+\frac{k\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^2+k\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}}{\delta_{n,T-k}^2}\right).$$

This result holds under either $k_{n,T}n^{-1/2}=o(1)$ or $k_{n,T}T^{-1/2}=o(1)$. Next, using (B.17) alongside Lemmas B.3(ii), (iii), (xi), and (xii), we can also straightforwardly demonstrate that

$$\left|\mathrm{E}\left(\frac{1}{n(T-k)}\sum_{j=1}^{n}\left(\frac{1}{T-k}\hat{\boldsymbol{\xi}}_{(k,k)}-\mathbf{X}_j^{(k,k)\prime}\mathbf{P}_{T-k}\mathbf{e}_j^{(k)}\right)'\hat{\mathbf{Q}}_{(k,k)}^{-1}\frac{1}{n}\sum_{i=1}^{n}\ddot{X}_{i,T}(k,k)\,\bar{e}_{i,T}(k)'\right)\right|$$
$$\leq C\left(\frac{k}{\delta_{n,T-k}^2(T-k)^{1/2}}+\frac{k\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}}{\delta_{n,T-k}^2}\right),$$

and

$$\left|\mathrm{E}\left(\frac{1}{n(T-k)}\sum_{j=1}^{n}\left(\mathbf{X}_j^{(k,k)\prime}\mathbf{M}_{T-k}\mathbf{e}_j^{(k)}+\frac{1}{T-k}\hat{\boldsymbol{\xi}}_{(k,k)}\right)'\hat{\mathbf{Q}}_{(k,k)}^{-1}\frac{1}{n}\sum_{i=1}^{n}\ddot{X}_{i,T}(k,k)\,\bar{s}_{i,T}(k,k)'\right)\right|$$
$$\leq C\left(\frac{k\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^2}{\delta_{n,T-k}^2}+\frac{k\|\boldsymbol{\alpha}(k)-\boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}}{\delta_{n,T-k}^2}\right)$$

by application of Hölder's inequality. These results are sufficient to establish Theorem 3.3.

B.3.2. Asymptotic Unbiasedness of QFR Estimator.

THEOREM B.10. *Let $y_{i,t}$ be generated according to (1). Under Assumptions 1, 2, and 3,*

$$\lim_{n,T\to\infty}\max_{1\leq k\leq k_{n,T}}\left|\frac{\mathrm{tr}\left(\Phi\left[\mathrm{E}\left(\hat{B}_{n,T}(k)\right)-\left(1-\frac{p}{T}\right)\Sigma\right]\right)}{\mathrm{tr}\left(\Phi B_{n,T}(k)\right)}-1\right|=0,$$

for all positive semidefinite $m\times m$ matrices Φ satisfying $\|\Phi\|=1$.

Theorem B.10 follows straightforwardly from Theorem B.4 and Lemma B.7, noting that $B_{n,T}(k)=L_{n,T}(k)-p^2\left(T-k_{n,T}\right)^{-2}\zeta(k)$ and $\tilde{R}(k)=\hat{R}(k)+p^2\left(T-k_{n,T}\right)^{-2}\hat{\zeta}(k)$.

B.3.3. Proof of Theorem 3.4 (*Asymptotic Efficiency*). The proof of Theorem 3.4 requires analogous versions of Lemmas B.8 and B.9 to hold for the bias-corrected forecast. Using the bound given in (B.17) and the expression for $\tilde{u}_{i,T+1}(k,l)$ given in (B.35), we can follow the same steps as given in the proof to Lemma B.9 to establish that

$$\mathrm{E}\left\|\frac{1}{n}\sum_{i=1}^{n}\left(\tilde{u}_{i,T+1}(k_2,k_2)-\tilde{u}_{i,T+1}(k_1,k_1)\right)\right\|_\Phi^{2q}\leq C_q\left(\left|\frac{k_2-k_1}{\delta_{n,T}^2}\right|^q+|l(k_2)-l(k_1)|^q\right)$$
(B.36)

for all $q > 1$ and k_1 and k_2 satisfying $1 \leq k_1, k_2 \leq k_{n,T}$. Next, since $\hat{B}_{n,T}(k) = \hat{L}_{n,T}(k) - p^2 \left(T - k_{n,T}\right)^{-2} \hat{\zeta}(k)$ and $B_{n,T}(k) = L_{n,T}(k) - p^2 \left(T - k_{n,T}\right)^{-2} \zeta(k)$, analogous versions of Lemma B.8(i) and (ii) hold for the BCLS case, specifically

$$\mathrm{E}\left(\left|\mathrm{tr}\left(\Phi \hat{B}_{n,T}(k)\right) - \mathrm{tr}\left(\Phi B_{n,T}(k)\right) - \mathrm{tr}\left(\Phi \hat{B}_{n,T}\left(k_{n,T}^{**}\right)\right) + \mathrm{tr}\left(\Phi B_{n,T}\left(k_{n,T}^{**}\right)\right)\right|^q\right) \tag{B.37}$$
$$\leq C_q \left(\frac{k^q + k_{n,T}^{**q}}{\delta_{n,T}^{2q} \left(T - k_{n,T}\right)^{\frac{q}{2}}} + \frac{k^q + k_{n,T}^{**q}}{n^{\frac{q}{2}} \left(T - k_{n,T}\right)^{\frac{3q}{2}}} + \frac{\|\boldsymbol{\alpha}(k) - \boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{2q} + \|\boldsymbol{\alpha}(k^*) - \boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{2q}}{\delta_{n,T}^{q}} + \frac{\|\boldsymbol{\alpha}(k) - \boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{q} + \|\boldsymbol{\alpha}(k^*) - \boldsymbol{\alpha}\|_{\Gamma\otimes\Phi}^{q}}{\delta_{n,T}^{q}} \right)$$

and

$$\sum_{k=1}^{k_{n,T}} \delta_{n,T}^{2r} \mathrm{tr}^r \left(\Phi B_{n,T}(k)\right) \mathrm{E}\left(\left| \frac{\mathrm{tr}\left(\Phi \hat{B}_{n,T}(k)\right) - \mathrm{tr}\left(\Phi B_{n,T}(k)\right) - \mathrm{tr}\left(\Phi \hat{B}_{n,T}\left(k_{n,T}^{**}\right)\right) + \mathrm{tr}\left(\Phi B_{n,T}\left(k_{n,T}^{**}\right)\right)}{\mathrm{tr}\left(\Phi B_{n,T}(k)\right)} \right|^q\right) = o(1) \tag{B.38}$$

for all $r > 1$ and q such that $q > 2(r+1)$. Theorem 3.4 then follows by applying the same arguments used to establish Theorem 3.2, using (B.36), (B.37) and (B.38), Theorem 3.3, and Lemma B.7.